JMIR Formative Research

Impact Factor (2022): 2.2
Volume 4 (2020), Issue 10 ISSN 2561-326X Editor in Chief: Gunther Eysenbach, MD, MPH, FACMI

Contents

Original Papers

Original Paper

Machine Learning–Based Screening of Healthy Meals From Image Analysis: System Development and Pilot Study

Kyoko Sudo[1*], DPhil; Kazuhiko Murasaki[2*], DPhil; Tetsuya Kinebuchi[2], MSc; Shigeko Kimura[3], DPhil; Kayo Waki[4], DPhil

[1]Department of Information Sciences, Toho University, Chiba, Japan

[2]NTT Media Intelligence Laboratories, Yokosuka, Japan

[3]Department of Ubiquitous Health Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

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

[*]these authors contributed equally

Corresponding Author:
Kyoko Sudo, DPhil
Department of Information Sciences
Toho University
Miyama 2-2-1, Funabashi
Chiba, 274-8510
Japan
Phone: 81 47 472 8064
Email: kyoko.sudo@sci.toho-u.ac.jp

Abstract

Background: Recent research has led to the development of many information technology–supported systems for health care control, including systems estimating nutrition from images of meals. Systems that capture data about eating and exercise are useful for people with diabetes as well as for people who are simply on a diet. Continuous monitoring is key to effective dietary control, requiring systems that are simple to use and motivate users to pay attention to their meals. Unfortunately, most current systems are complex or fail to motivate. Such systems require some manual inputs such as selection of an icon or image, or by inputting the category of the user's food. The nutrition information fed back to users is not especially helpful, as only the estimated detailed nutritional values contained in the meal are typically provided.

Objective: In this paper, we introduce healthiness of meals as a more useful and meaningful general standard, and present a novel algorithm that can estimate healthiness from meal images without requiring manual inputs.

Methods: We propose a system that estimates meal healthiness using a deep neural network that extracts features and a ranking network that learns the relationship between the degrees of healthiness of a meal using a dataset prepared by a human dietary expert. First, we examined whether a registered dietitian can judge the healthiness of meals solely by viewing meal images using a small dataset (100 meals). We then generated ranking data based on comparisons of sets of meal images (850 meals) by a registered dietitian's viewing meal images and trained a ranking network. Finally, we estimated each meal's healthiness score to detect unhealthy meals.

Results: The ranking estimated by the proposed network and the ranking of healthiness based on the dietitian's judgment were correlated (correlation coefficient 0.72). In addition, extracting network features through pretraining with a publicly available large meal dataset enabled overcoming the limited availability of specific healthiness data.

Conclusions: We have presented an image-based system that can rank meals in terms of the overall healthiness of the dishes constituting the meal. The ranking obtained by the proposed method showed a good correlation to nutritional value–based ranking by a dietitian. We then proposed a network that allows conditions that are important for judging the meal image, extracting features that eliminate background information and are independent of location. Under these conditions, the experimental results showed that our network achieves higher accuracy of healthiness ranking estimation than the conventional image ranking method. The results of this experiment in detecting unhealthy meals suggest that our system can be used to assist health care workers in establishing meal plans for patients with diabetes who need advice in choosing healthy meals.

(JMIR Form Res 2020;4(10):e18507) doi:10.2196/18507

KEYWORDS

meal images; healthiness; deep neural network; nutrition; medical informatics; diet; neural network

Introduction

Recently, many information technology–supported systems for health care have been developed, including systems using image-based dietary assessment for obesity and diabetes management [1]. A survey of popular nutrition-related mobile apps demonstrated that there is clear interest for diet monitoring and recommendation using mobile apps [2]. With the inclusion of a camera feature, mobile devices are increasingly being used for image-based dietary assessment. One of these systems is DialBetics [3], which is assisted by the food image recognition app FoodLog [4] to input photos of meals by semiautomatically limiting the dish-selection area. Herein, we refer to a "meal" as the dish or dishes eaten by a person in a single sitting; thus, a single dish or a multiple-dish menu may constitute a meal. A narrative review of this system indicated that most patients actively used the dietary evaluation module, and each meal's nutritional balance sent to the patients helped them to modify their diet. However, the image processing was used only to assist inputting meal photos and identifying the name of meals; the meals' total energy, macronutrients, dietary fiber, and salt were calculated by dietitians from the photos.

However, two problems can cause people to stop using the system. The first problem is that most current technologies require user action to achieve meal image recognition [4,5]. Estimating nutrition automatically from only inputting meal images by users is expected to be an important function of information technology–supported systems for health care control. However, this has not been achieved by most of these systems, including DialBetics [3]. For instance, with DialBetics, the user may need to manipulate images so that only one dish is included, or may need to identify the food area within the broader plate area. The user generally must select the food category among those suggested by the system, and when some foods do not match any of the available categories, users must register a new category in the system. These fairly difficult tasks lead some users to discontinue use of the system [1,3].

The second problem is the disconnect between the system's output and the user's understanding. The interviews with users of Dialbetics showed that some users consider the advice obtained from the dietitian to be too long and redundant [3]. It is difficult for users to know whether their meals are good for them based on a simple listing, however detailed, of nutritional values. Therefore, a system is needed that offers immediate feedback allowing users to understand the nutritional implications of what they are eating and motivates them to become interested in eating better meals.

It seems to be a safe assumption that a "meal image," a single image of all of the dishes in a meal, contains some visual clues that permit estimation of the healthiness of the meal. Using meal images, we aim to provide a simple interface between the user and the system, with clear feedback that motivates users to continue to care about their meals, and a way to screen those who need the advice of a registered dietitian to improve their meals. A previous study of a system that allows sharing meal feedback from other users based on images [6] showed that the feedback, even from users rather than experts, is effective in improving diet. Our system is based on a similar concept. The feedback is not detailed numeric nutrition data but rather a more intuitive rating that can be easily understood by users. We targeted health promotion and diabetes, since the basic conditions of ideal meals in these contexts can be shared, including adequate energy and balanced nutrition, avoidance of salt, and obtaining abundant dietary fiber. We also designed the system to assess one meal eaten at a time, since the guidelines of a healthy diet for patients with diabetes in Japan suggest having three meals a day, and ensuring a nutritionally balanced and equalized portion in each meal as much as possible [7].

Many food image recognition systems have been proposed in recent years [5,8-20]. The most popular approach uses general object recognition technologies that sort foods into categories. The development of machine learning and a large database have accelerated growth in the number of food categories that can be recognized in recent years [17-19]. One approach to nutrition estimation is category-based estimation; that is, recognizing the category of food and displaying the nutritional value of that category. This approach does not assess the amount of food, and value outputs assume that one dish contains a regular portion or requires users to select among a lineup of values for multiple people or photos of different portion sizes [14]. To estimate the amount of food in a given dish, each food area must be segmented and the volume of each segment estimated. Other recent approaches include one based on a convolutional neural network (CNN) that trains a model using large sets of food images and their nutritional values. To estimate nutrition more accurately, one approach simultaneously posits the name of the meal and the nutrition of that food [9,11], and another estimates the ingredients in each dish and their proportions/weights from which it calculates nutritional value [15,20]. Clearly, these food recognition systems based on machine learning require a large amount of training data. Recent research has involved food recognition with an original food dataset containing hundreds of food categories [17-19]. Nevertheless, given the extremely high number of different categories of food in the world, covering all of them with such a model is virtually impossible. Furthermore, home cooking often involves dishes that are hard to assign to an appropriate category name, making it difficult to conduct training. To identify food categories, some approaches recognize foods by their ingredients [12,13]. These approaches are valid when the appearance of the ingredients changes minimally during cooking. However, in most cases, the appearance will change greatly after cooking, making the scope of this approach rather limited.

In contrast to these conventional methods that estimate nutrition by recognizing the category of foods or ingredients, we propose a novel approach of predicting whether or not a meal is healthy by viewing the total meal image. Our approach avoids the highly difficult food image recognition task that determines the absolute

values of foods. Instead, our healthiness estimation method uses a ranking network and the ranking data generated by comparing many pairs of meal images. We fuse a recognition network, trained with a food dataset and with masks of food areas, with a ranking network. This approach allows for extracting food features from meal images that contain multiple dishes, enabling rendering an accurate judgement about the entire meal. Here, we report the results of a pilot study to highlight the pipeline of our proposed system from generating the ranking data of healthiness of meals to unhealthy meal detection.

The machine-learning algorithms that can be used for comparisons among data are collectively referred to as machine-learning rank algorithms (MLRAs). Several MLRAs based on support vector machine (SVM) or CNN algorithms have been proposed. CNN is known to offer high performance in recognition tasks, and CNN-based rank algorithms have been applied to estimate various attributes of images. One study demonstrated the estimation of town attributes from urban images of landscapes [21]. We employed the same MLRA, and trained its CNN using meal images manually ranked in terms of healthiness. One difference between learning the rank of landscape images and learning the rank of meal images is that the healthiness of meals should be estimated solely from the foods or ingredients. Accordingly, meals that contain the same ingredients but are photographed on different plates should be estimated to have the same healthiness.

It is also important to prevent the network from learning relationships between healthiness and factors other than the food itself. Toward this end, our ranking network was structured to estimate rank using only the food area by learning using spatially selected areas, implemented using a mask of the region of interest in the training phase. Our network was then trained by many ordered image pairs, according to the ranking method of Dubey et al [21] or TrueSkill [22].

Although a network could be trained online, we adopted an offline approach for learning. We first established a ranking dataset, and then input pairs of images from the dataset when training the network. We used this approach because the ranking data are needed to calculate the deviation value of the ranking score, which is the output of our system that intuitively expresses the healthiness of meals. Our database was annotated not by crowdsourcing but rather by an expert (a registered dietitian). Establishing the ranking dataset in advance is helpful for the expert, which allows them to work at any convenient time and can redo the work if warranted. Once establishing the ranking data, we simply generated the sets of pairs of rankings from this dataset.

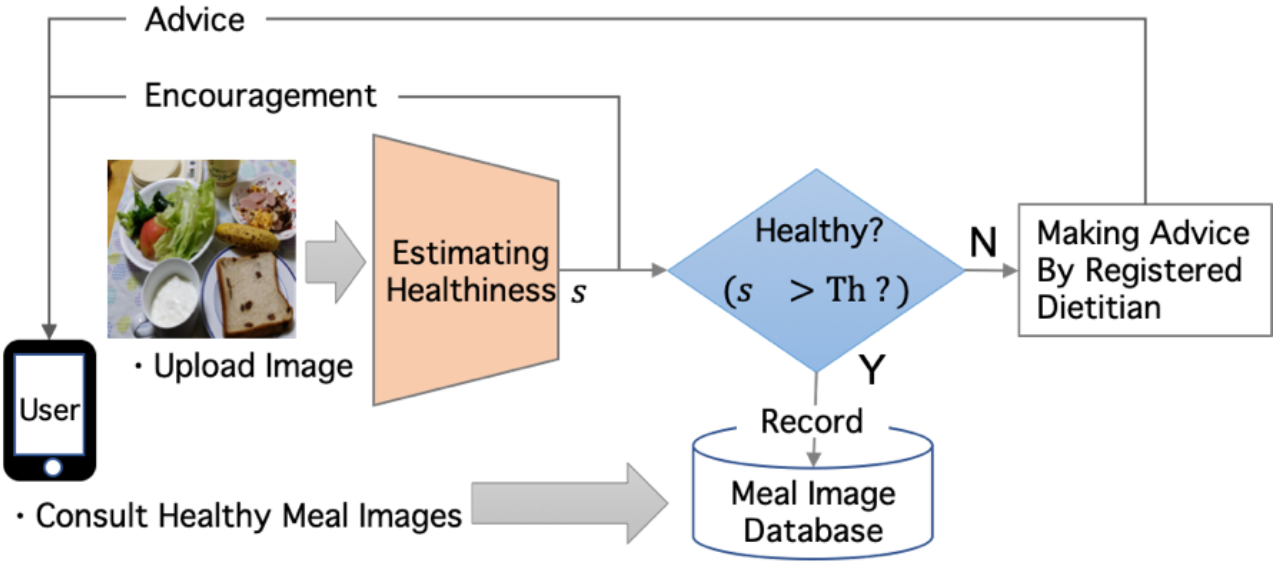
Rank-SVM [23,24], an MLRA-based approach, is also used for image retrieval. In the model proposed by Joachims [23], the rank of images was learned using a ranking dataset that had a few hundred images per category. In view of the limited size of our dataset, there is a possibility that Rank-SVM would perform better than CNN; therefore, we compared the performance of ranking score–based CNN to Rank-SVM in this experiment to optimize the system.

Methods

Workflow

We propose an image-based system that can rank meals in terms of the overall healthiness of the dishes constituting the meal. First, we generated a database of meal images ranked by a registered dietitian viewing the images. We then constructed a network that maintains conditions important for judging the meal image, while extracting features that eliminate background information and those that are independent of location. The output of the network, *healthiness*, and the dietitian's judgment were expected to be related with a high correlation coefficient. The workflow of our system is schematically shown in Figure 1.

Figure 1. Workflow of the screening system and the image database of healthy meals for those who need the advice of a registered dietitian to improve their meals.



Our system extracts the *healthiness score*, which is calculated as the deviation from average healthiness, so that users can know whether or not their meal is a healthy choice within the distribution of healthiness of meals by assessing how far they are from the average healthiness. In Figure 1, *s* is the healthiness score estimated by the proposed method and *Th* is the threshold of healthiness in screening dishes. When the healthiness score is high, the system sends the user the score and a message encouraging them to continue consuming healthy meals and using the system. The system further allows the users to check their records of the meal images with their healthiness scores; this function serves as a reference for the users to choose healthier meals. When the healthiness score is low ($s<Th$), a registered dietitian intervenes to help the user modify their diet. These users can consult the database for healthy meal images to improve their meals. Because the users obtain feedback immediately after they record their meals, they can change to a healthier meal if their initial choice achieves a low score.

Framework Overview

Our image-based meal rating system performs the following processes: (1) examines whether a registered dietitian can judge the healthiness of meals solely by viewing meal images in a small dataset (100 meals); (2) generates ranking data based on comparisons of sets of meal images (850 meals) by a registered dietitian viewing meal images; (3) trains a network (a feature-extraction subnetwork, pretrained by a food dataset before being trained by our ranking dataset, and a ranking estimation subnetwork) based on the ranking to estimate a health metric; and (4) estimates each meal's score, and the domain adaptability of estimating each meal's score to finally detect unhealthy meals based on the health metric.

Ranking Meal Healthiness

Ground Truth and Dietitian's Subjective Evaluation

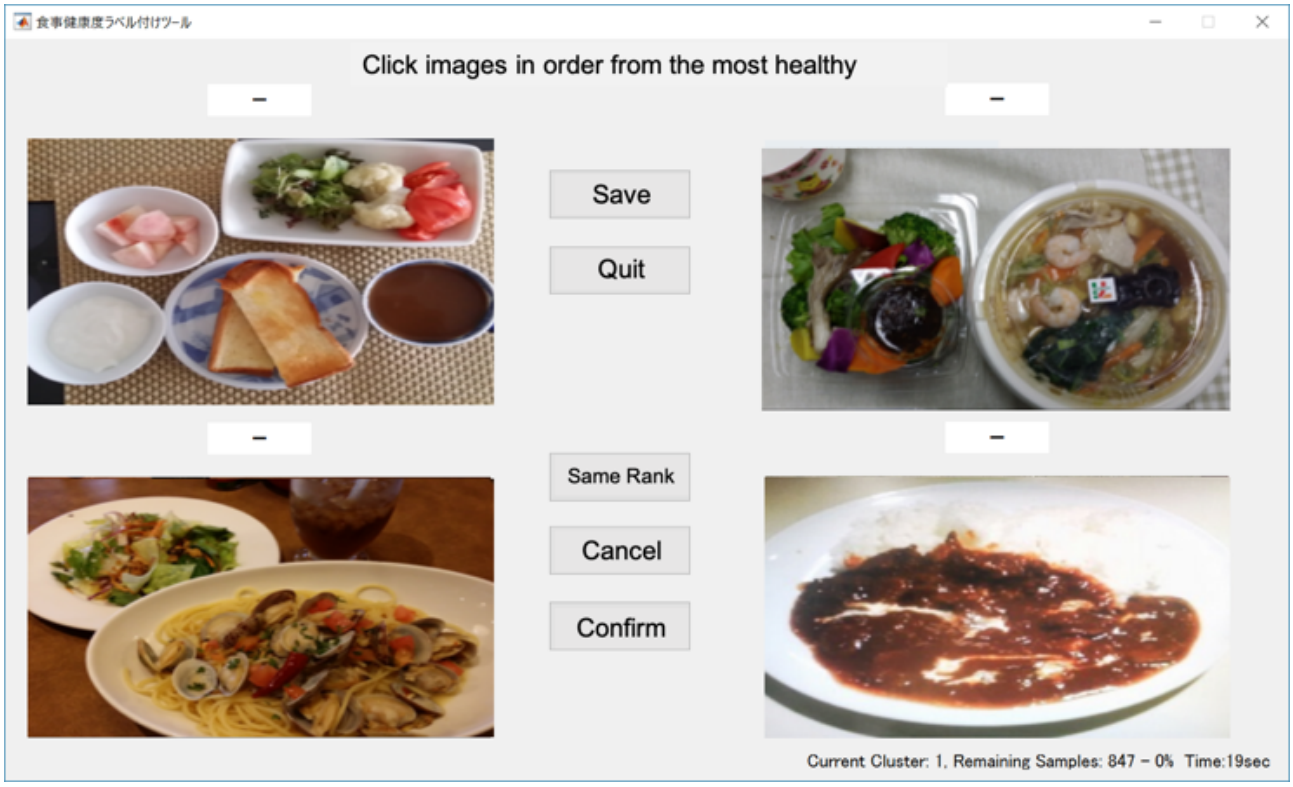
Ground truth consists of the sets of images and their associated ranks. The ground truth can be developed through cooking sample meals. The registered dietitian cooks the sample meals using ingredients with measured nutritional values to allow for accurate calculation of the total nutritional values using proven nutrition formulas [7]. The meal is then photographed and the set of nutritional values is ranked, resulting in a set of images and ranks.

This methodology is resource-intensive, and in practical situations leads to small datasets due to resource constraints. Therefore, we required larger datasets to train the network. Accordingly, we expanded the ground truth set using experts to perform subjective assessments of the healthiness rank based on images alone. To establish the validity of this approach, we verified that the expert, a registered dietitian, can appropriately judge the healthiness rank by viewing a meal image. We asked the dietitian to examine the images in the ground truth set and to rank them according to their nutritional value. The dietitian who judged the meal images was different from the dietitian who cooked the sample meals for the ground truth dataset and photographed them. We then appraised the relationship between the rank given by the dietitian viewing images of the meals and the ground truth rank based on nutritional measurements. The ground truth rank based on nutritional measurements was calculated according to the standard values of food composition in Japan [25,26]. Briefly, total energy and the energy ratio (ie, the ratio of protein, fat and carbohydrate), and the supplementary items (appropriateness of salt, and the amount vegetables, beans, and foods rich in dietary fiber such as mushrooms) are the items required for the calculation.

Generating Ranking Data

To generate a dataset containing the images and the healthiness metric implicit in each image, we used a custom app displaying a set of multiple meal images that were ranked by dietitians according to healthiness from "best" to "worst." A dietitian was told that ranking should indicate whether the dishes could appropriately constitute an entire meal, and was also told to make the judgment as precisely as possible from only viewing the image of each meal. If it was difficult to differentiate between meals, the same rank assigned to multiple meals was acceptable. To reduce the dietitian's workload, our app displays only 4 images at each step. The app interface is shown in Figure 2. For this study, we used the meal image database of patients with type 2 diabetes.

Figure 2. App interface to generate the ranking data. A set of 4 meal images is presented, and the registered dieticians rank those images from the 1st to 4th.



The ranking algorithm is shown in Multimedia Appendix 1, in which N_c=4. We generated ranking data by repeatedly ranking sets of multiple meal images consisting of one pivot image and other images, according to step 17 in the algorithm, conducted by the dietitian. To generate a larger ranking dataset in the future, the TrueSkill [22] algorithm can be used, which considers the distribution of annotation that merges the rank sets annotated by multiple registered dietitians.

Proposed Network and Train Ranking

We trained the network to output the right ranking order of randomly selected pairs of images using the approach proposed by Dubey et al [21]. This approach was originally proposed for training urban images, where ranking was associated with safety, features were extracted from the whole image, and the score was output as a scalar value. Therefore, we had to adapt this algorithm to meal images.

The healthiness of the meals should be estimated solely from the comprising foods or ingredients. Therefore, meals that contain the same ingredients but were photographed on different backgrounds or plates should be estimated to have the same healthiness. To assure this continuity, we modified the ranking layer so that pixel features were only those included in the region of interest, eliminating any possible impact of food placement or background such as tables. In the training phase, we prevented the network from learning relationships between healthiness and factors other than the food itself by using a mask to indicate the region of food. The mask simply used a 1 value in the food area and a 0 value elsewhere, so that pixel-by-pixel multiplication of the original image and the mask resulted in an image where the food area is the same as the original image and any part of the image outside of the food area is given a value of 0. The mask was generated manually, outside the proposed system. The mask is applied only when the data are used for training; once the network is trained, masking is no longer necessary.

The network to predict healthiness was trained using the expanded ground truth ranking data generated by the registered dietitians. Pairs of images and their relative rankings were input in the training process, and all pairs were labeled to indicate which is healthier than the other. Duplicate networks were used to predict healthiness. Their outputs were used to calculate the loss, defined as:

(1)

(2)

where x_i, x_j is the pair of images i, j for training, $\in$ is the set of all pairs of labeled images, and $f(x)$ is the estimated healthiness of image x.  and  are the ground truth healthiness, and equation (2) is the relation between the ground truth order of images x_i and x_j.

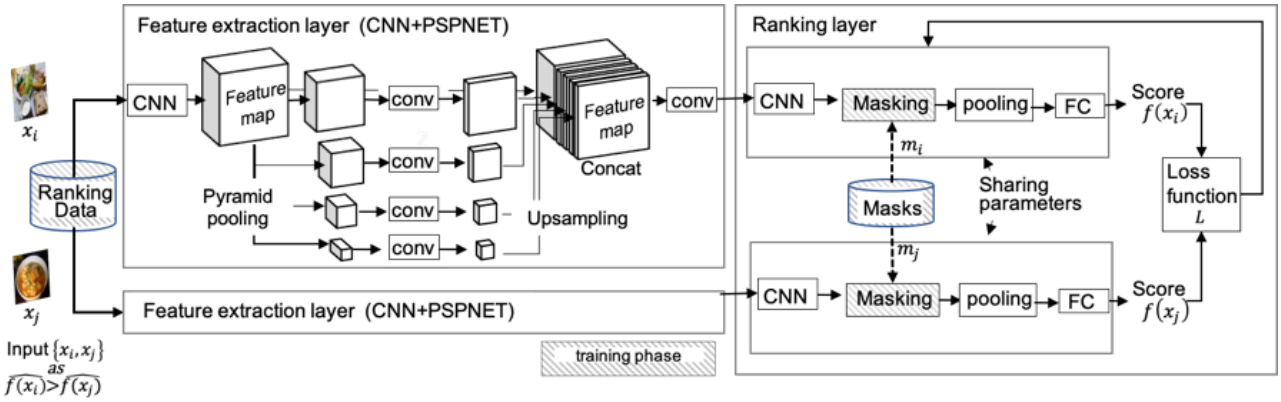
We assign i and j for a pair of images as they satisfy  . Since the value of the loss function of equation (1) becomes smaller when the order of the estimated ranks $f(x_i)$ and $f(x_j)$ is $f(x_i)>f(x_j)$ in the condition that the ground truth order relation between images i and j is  , the ranking predictor is trained so as to minimize the loss for all data. When $f(x_i)>f(x_j)$, it is true, and L

will give a value close to 0. When the order of the estimated rank $f(x_i)$ and $f(x_j)$ is $f(x_i)<f(x_j)$, it is false, and L will give a large value.

As the network for feature extraction, we used the same architecture as used in the pyramid scene parsing network (PSPNET) [27]. This provides pixel-level category prediction. To obtain the explicit relation between the feature vector and the feature of the local region of the food, the feature extraction module was trained in advance. It is possible to train this module by optimizing both ranking and food class; however, a dataset that has annotation of both rank and food class is not available. Therefore, we adopted a serial approach by pretraining the feature extraction layer with a large food class dataset and then connecting it to the ranking layer, followed by train ranking in an end-to-end manner.

We used the UEC FOOD-100 dataset [5,19], which includes 100 food categories, for pretraining the feature extraction layer. The network was pretrained to output the correct category of food. PSPNET [27] in the feature extraction layer was pretrained to the output food category by pixel. The entire architecture is shown in Figure 3.

Figure 3. Proposed system for training ranking data of meals. The feature extraction layer consists of the convolutional neural network (CNN) and the same network as that of the pyramid scene parsing network (PSPNet) [27], which outputs pixel-wise feature maps. The ranking layer module is for estimating the scores. FC: fully convolutional layer.



Since the assertion of healthiness by the network has no meaning per se, we defined the deviation as the healthiness value, which was calculated as the distance from the mean value of the network output normalized by the variance. We call this variance the *healthiness score*, which is different from the *rank*.

Experiments

Verifying the Accuracy of a Dietitian's Subjective Evaluation

The expert (registered dietitian) had to infer ingredients and foods from the images.

We then confirmed the relation between the ranking based on viewing images with rankings according to the measured nutritional value in advance using the registered dietitian's image-based rank to create an expanded ground truth set in our experiments.

Verifying the Accuracy of the Meal Rating Model

We conducted our experiment taking into consideration any errors in pair comparison and ranking estimation, unhealthy meal detection, and domain adaptability of meal images.

We used an original meal image database of patients with type 2 diabetes (see the General Ranking Data section below for the detailed process of ranking annotation) and the UEC Food Dataset [19]. We used 90% of the ranking data by the dietitian for training the proposed CNN-based ranking estimation system, and the rest of the ranking data were used for pair comparison/rank estimation and unhealthy meal detection.

For pretraining of the CNN and evaluation of the domain adaptability of the ranking estimate, we used the images in UEC Food Dataset.

Pair Comparison and Rank Estimation

We evaluated the accuracy of ranking healthiness under different conditions: (1) with and without pretraining the feature extraction layer, (2) with and without using a mask generated by semantic segmentation, (3) with CNN+FC (using a fully convolutional layer connecting CNN and Rank-SVM) or Rank-SVM as the output layer in the training and test phase, and (4) with rank-based CNN [21] using a mask.

We conducted experiments to examine the contribution of pretraining the CNN for feature extraction, an end-to-end structure using a CNN for the output layer compared to using Rank-SVM, and to examine the performance of our method compared to the original rank-based CNN proposed by Dubey et al [21]. Since the original rank-based CNN (RSS-CNN) does not have the structure of using a mask that indicates the region of food, we masked the region other than food in the input image by embedding 0 values to compare with our method in the same condition. We compared two networks: one based on RSS-CNN and the other based on Rank-SVM. The latter outputs the rank of healthiness by Rank-SVM using the CNN feature, which is generated by the same trained CNN as the network constructed from the RSS-CNN–based network.

The evaluation indices were the error rate of comparison E_p and the average error of order E_o determined as follows.

When the order relation between the estimated scores s_i and s_j of a pair of test images I_i and I_j are the same as the order relation between the pair of scores $\hat{s}_i$ and $\hat{s}_j$ assigned by the registered dietitian to I_i and I_j, the score estimation result for the pair i and j is counted as true, $C(i,j)=1$.

The accuracy of the ranking of pairs was calculated for all pairs of test images. The ratio of counts that proved true is then taken to be the error rate of comparison E_p of the pairs.

Where Ω is the set of all other pairs of test samples (I_i, I_j), $(i, j \in \{I_1, I_2,\ldots, I_N\})$ and N is the number of test samples.

We generated two rankings by sorting the test samples according to two scores: those estimated by the proposed method and those assigned by the registered dietitian. We then compared the order o_i of I_i to the order $\hat{o}_i$ by the registered dietitian, and calculated the error between o_i and $\hat{o}_i$ as $e_i = \mid o_i - \hat{o}_i \mid$; o_i is obtained by sorting the images according to the estimated ranks.

The average error of order E_o is taken to be the mean of e_i; Ω is the set of test samples $\{I_1, I_2,\ldots, I_N\}$ and N is the number of test samples.

Unhealthy Meal Detection

The proposed system can be used to detect unhealthy meals. It can automatically identify which patients tend to select unhealthy meals and trigger oversight by health experts. To confirm this, we conducted an experiment on unhealthy meal detection.

Domain Adaptability of the Ranking Estimate

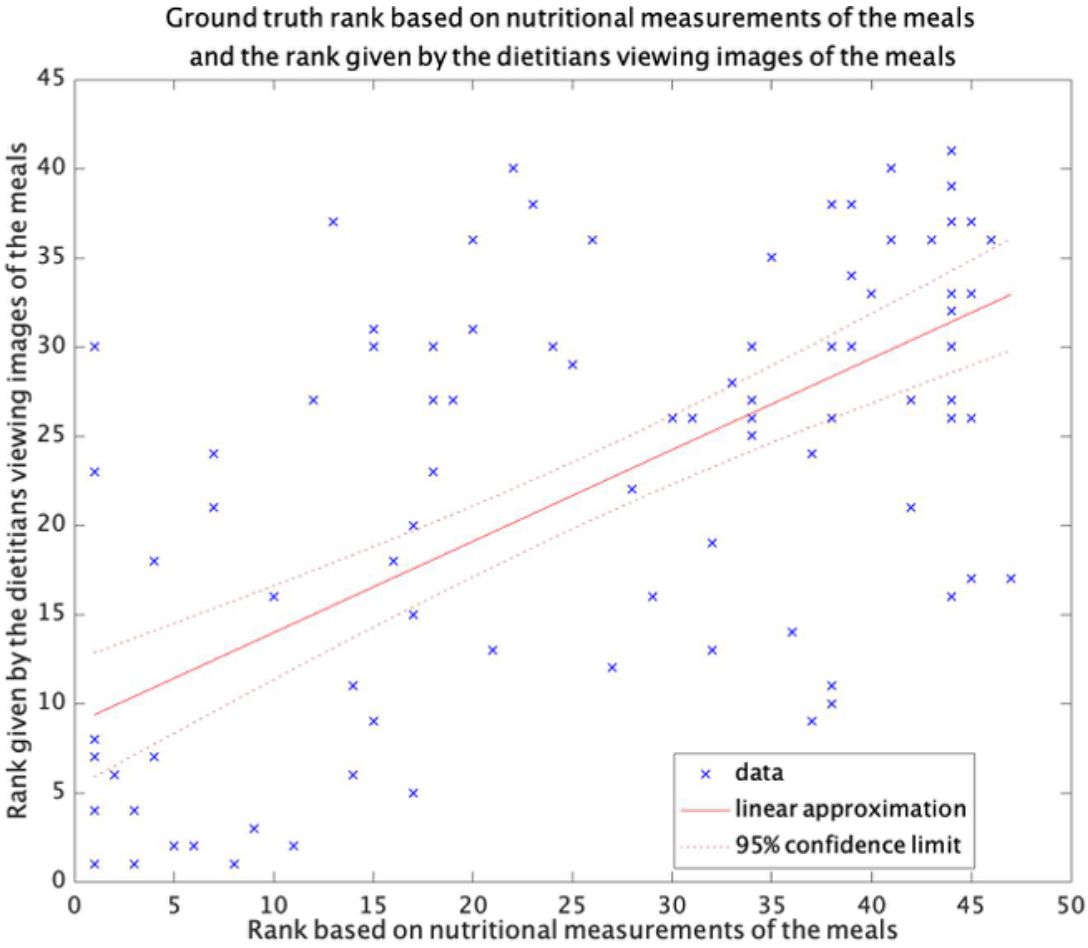
To test the domain adaptability of our method, we estimated the healthiness scores of the images of a publicly available database, UEC Food Dataset [5,19], whose domain is different from that of our training dataset.

Results

Dietitian Subjective Evaluation

Figure 4 shows the correlation between the rank judged based on viewing images and the ground truth rank. With a correlation coefficient of 0.73, we confirmed that ranking based on viewing images correlates with ranking based on the measured nutritional value. The Bland-Altman plot further confirmed that there is no fixed bias or proportional bias (Multimedia Appendix 2). Accordingly, we used a registered dietitian's image-based rank to create an expanded ground truth set in our experiments.

Figure 4. Ground truth rank based on nutritional measurements of the meals and the rank given by the dietitian viewing images of the meals.



Pair Comparison and Rank Estimation

Table 1 shows the results of the accuracy achieved in ranking healthiness under different conditions. The result of the proposed method with masking, pretraining, and the CNN was better than that of the network without any module related to feature extraction.

Table 1. Evaluation of the error rate of healthiness between methods.

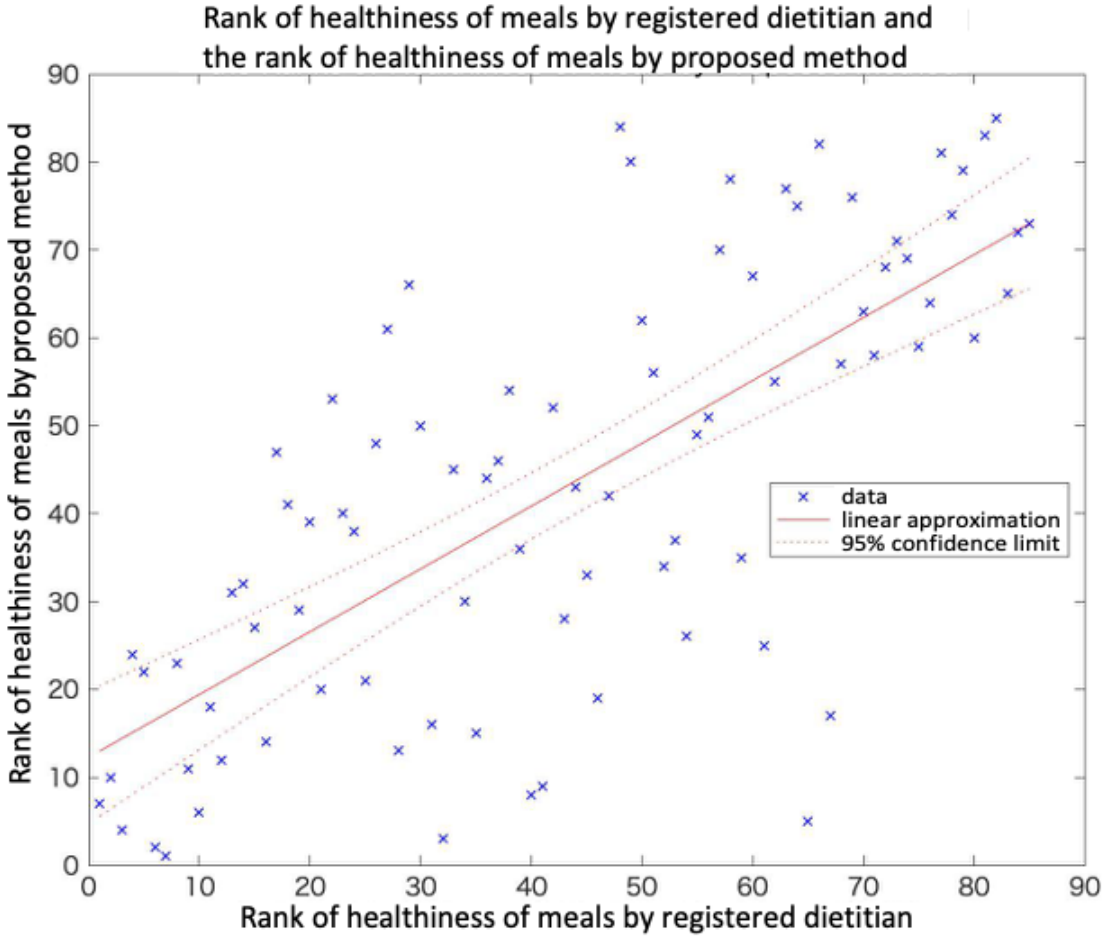
Method	Conditions for training the model			Error rate of the rank (E_o)	Error rate of the pairwise rank (E_p)
	Masking	Pretraining	Ranking method		
Proposed method	Yes	Yes	CNN[a]	13.94	16.40%
	Yes	No	CNN	14.59	17.16%
	No	Yes	CNN	16.6	19.5%
	Yes	Yes	Rank-SVM[b]	16.2	19.1%
Rank-based CNN [21]	Yes	No	CNN	15.44	18.15%

[a]CNN: convolutional neural network.

[b]SVM: support vector machine.

The relation between the ranking of healthiness based on the dietitian's judgment and the ranking estimated by the proposed method is shown in Figure 5. The rankings were normalized to have a mean of 0 and unit variance, and were then transformed to range from 0 to 100. The correlation coefficient between these rankings was 0.72. We also confirmed that there was no fixed bias or proportional bias by evaluation of the Bland-Altman plot (Multimedia Appendix 3).

Figure 5. Comparison of the normalized ranking by the registered dietitian and the ranking estimates output by the proposed method.



Unhealthy Meal Detection

Table 2 shows the results of the accuracy achieved in unhealthy meal detection by the proposed and rank-based CNN methods. Although the difference between methods was small, the proposed method with the full process was again superior.

Table 2. Evaluation of the accuracy of unhealthy meal detection.

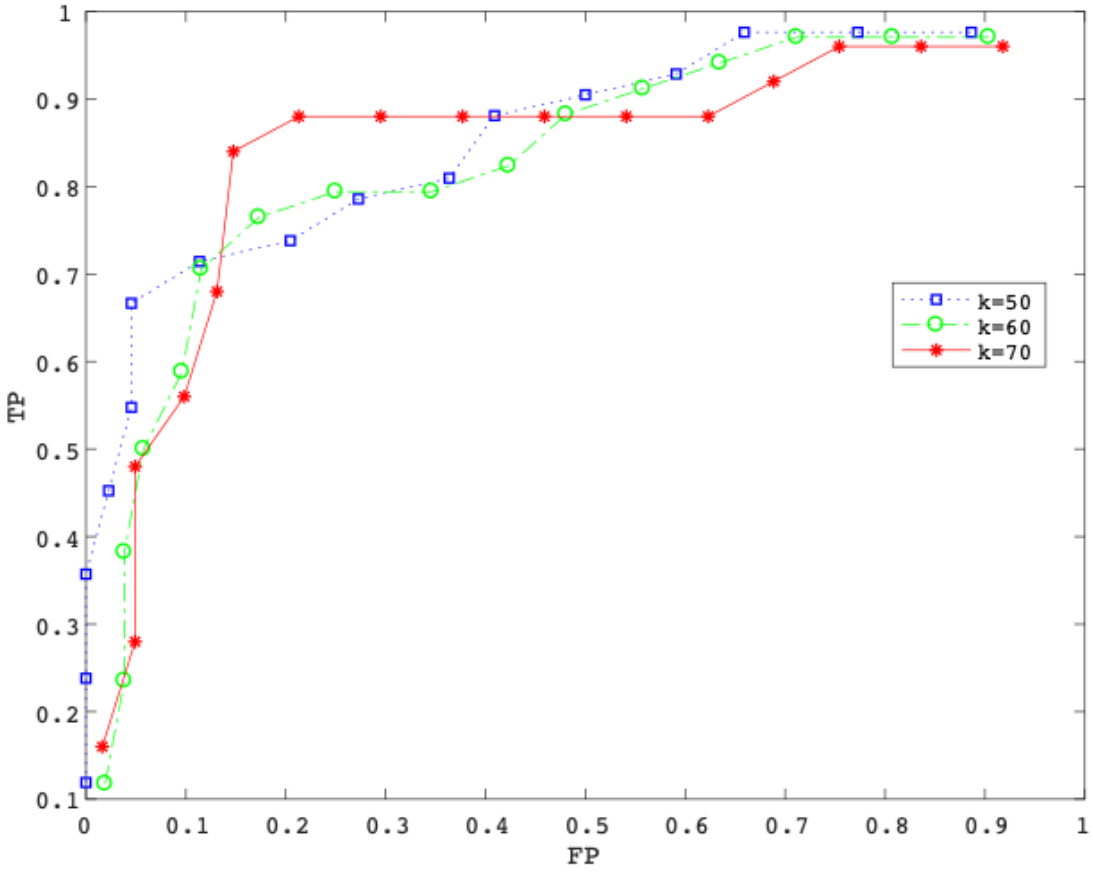
Method	Conditions for training the model			Accuracy of unhealthy meal detection
	Masking	Pretraining	Ranking method	
Proposed method	Yes	Yes	CNN[a]	76.5%
	Yes	No	CNN	73.9%
	No	Yes	CNN	72.5%
	Yes	Yes	Rank-SVM[b]	70.3%
Rank-based CNN [21]	Yes	No	CNN	72.86%

[a]CNN: convolutional neural network.

[b]SVM: support vector machine.

The curves in Figure 6 show the rate in the number of meals that were detected as unhealthy but are actually healthy in reality versus that of meals detected as unhealthy and are unhealthy in reality. Each curve shows the values when “unhealthy” was defined as a meal ranking in the lower *k*% (*k*=50, 60, 70) of all meals.

Figure 6. Receiver operating characteristic curves of unhealthy meal detection. The horizontal axis is the false positive rate (FP) and the vertical axis is the true positive rate (TP). Each curve shows the set of values of TP vs FP, when the definition of an unhealthy meal is a meal whose rank is lower than k% (k=50, 60, 70) of all meals.



Sample images and their healthiness values determined by the proposed method are shown in Multimedia Appendix 4, including images that have high or low scores, and examples of images with large error rates. These images show meals that were predicted to be very healthy yet deemed very unhealthy by the experts and vice versa.

Domain Adaptability of Ranking Estimate

Multimedia Appendix 5 shows sample meal images from UEC-Food Dataset whose domain is different from that used in our training dataset, and the respective healthiness scores estimated by the proposed method, including meals that have higher and lower scores.

Discussion

Principal Findings

Our experimental results show that there is a relation between the ranking of healthiness based on the dietitian's judgment and the ranking estimated by the proposed method (Figure 5), and it is possible that the CNN acquired something similar to human intuition. Multimedia Appendix 4 shows samples of meal images and the deviation values of healthiness as estimated by the proposed method. The meal images with high scores contained more dishes and more red or green colors than meal images with low scores. Although we cannot infer a cause from this finding, it is reasonable to suppose that that food color or the numbers of dishes in the image could have impacted the healthiness prediction.

In the case of images with large error rates (Multimedia Appendix 4), the red color of the raw meat or the yellow color of tempura in the left image may have induced the system to rate the dishes as healthy even though the dietitian judged them as unhealthy because of excessive calories. Conversely, the images below show meals that were predicted to be unhealthy but judged as healthy by the expert. The *yakisoba* dish in the left image consisted of stir-fried noodles with a lot of vegetables and meat; the dietitian judged this meal as healthy, but its color may have induced a negative prediction from the system. Since the logic of the prediction of healthiness by the network is not explicit, the impact of food color or number of dishes is uncertain. However, it appears that some implicit criteria formed by the dietitian had been transferred to the network model.

The images in Multimedia Appendix 5 show samples with scores that are relatively high or low in UEC Food Dataset, which uses a different domain from our original dataset in the training the model. Most of the meals with higher scores contained multiple dishes that have balanced, nutritious ingredients (ie, meat plus

vegetables). The meals with lower scores are mainly single-plate dishes with no vegetables and high carbohydrates.

Although the meals were cooked at different times and included slightly different ingredients, some visual clues are associated with healthiness, and we can assume that the proposed system uses these visual clues in assessing healthiness.

In this work, we trained the network to learn a model based on the dietitian's definition. It is possible to provide multiple indices in the future by establishing training data using our ranking GUI tool and training other models.

Based on pair comparison and rank estimation, we found that both masking and pretraining were effective methods to learn meal healthiness. The results of the end-to-end structure using a ranking layer were better than those obtained using a feature extraction layer and Rank-SVM. This suggests that the end-to-end approaches used by the ranking layer achieved better performance than feature-based extraction approaches used by Rank-SVM. The correlation coefficient of 0.72 between the rank of the proposed method and the rank given by the dietitian was not particularly high; however, this correlation and the error rate of pairwise comparison prediction of 16.4% (accuracy of 83.6%) are in line with previous work, including the accuracy of the original ranking method (ranking the safety of the city from its image) [21] of 73.5%, and a related study using machine learning–based calorie estimation from meal images that contain a single plate [20] that reported a correlation coefficient of the estimated calorie and the ground truth calorie of meal images of 0.78.

The result for unhealthy meal detection suggests that it is possible to set some appropriate thresholds that balance false positives and true positives. For example, when 60 meal images (2 meals from 30 users) are uploaded each day, by defining "unhealthy" as a meal ranking in the lower 50% of all meals, a threshold can be selected so that the lower 30 images are automatically detected with only a few healthy meals included. The assessment of domain adaptability of the ranking estimate suggests that our method has domain adaptability, so that meal images taken in various conditions (ie, at home or in restaurants) will be acceptable.

Since our meal image database consists of the photos of real meals of patients with type 2 diabetes, the scale of the database is not large, and the food categories are limited. In addition, the practically effective dataset for training the ranking model is even smaller since we allowed for tied ranks when the dietitian annotates the rank of the meal images. The number of dietitians giving a rank to part of the meal image database for training was also limited. Currently, we have data from two dietitians. In this work, we used the data from only one dietitian owing to the larger size of the dataset. Because of these limitations, the results of our experiments must be interpreted in light of the context of a pilot study. Generating a larger database with more categories of meals will help to improve the accuracy of ranking estimation.

We generated the model of ranking estimation using a machine-learning approach under the assumption that there is a relation between the appearance of meal images and the rank given by a dietitian. However, if a larger-scale dataset is available, there is a possibility that we will be able to classify meals into multiple categories for both healthy and unhealthy meals.

Conclusions

We have presented an image-based system that can rank meals in terms of the overall healthiness of the dishes constituting the meal. First, we showed that the ranking has good correlation to nutritional value–based ranking. We then proposed a network that allows conditions that are important for judging the meal image, while extracting features that eliminate background information and are independent of location. Under these conditions, the experimental results showed that our network achieves higher accuracy of healthiness ranking estimation than the conventional image ranking method. Although the size of the training dataset is not yet sufficiently large for a training-only solution, introduction of pretraining of the feature extraction network using the food dataset enables the system to produce estimated rankings with high correlation to the ranking of an expert.

The results of this experiment in detecting unhealthy meals suggest that our system can be used to assist health care workers in establishing meal plans for diabetic patients who need advice in choosing healthy meals. Future work will include creating a larger dataset using the ranking data of multiple registered dietitians and improving the accuracy of inference.

Acknowledgments

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant (16K09163).

Conflicts of Interest

SK and KW are members of the Department of Ubiquitous Health Informatics, which was engaged in a cooperative program between the University of Tokyo and NTT DOCOMO.

Multimedia Appendix 1
Algorithm for generating the ranking dataset.
[DOCX File , 14 KB - formative_v4i10e18507_app1.docx]

Multimedia Appendix 2
Bland-Altman plot of the ranks of healthiness of meals by the dietitian and the proposed method.
[PNG File , 116 KB - formative_v4i10e18507_app2.png]

Multimedia Appendix 3
Bland-Altman plot of the mean value and the difference between the ground truth rank based on nutritional measurements and the rank given by the dietitian viewing images of the meals.
[PNG File , 133 KB - formative_v4i10e18507_app3.png]

Multimedia Appendix 4
Samples of meal images and the deviation values of healthiness as estimated by the proposed method (top), and samples of meal images with large error rates (bottom).
[DOCX File , 785 KB - formative_v4i10e18507_app4.docx]

Multimedia Appendix 5
Sample of meal images from UEC-Food Dataset and their healthiness scores estimated by the proposed method for meals with (a) higher scores and (b) lower scores.
[DOCX File , 370 KB - formative_v4i10e18507_app5.docx]

References

1. Rollo ME, Aguiar EJ, Williams RL, Wynne K, Kriss M, Callister R, et al. eHealth technologies to support nutrition and physical activity behaviors in diabetes self-management. Diabetes Metab Syndr Obes 2016;9:381-390. [doi: 10.2147/DMSO.S95247] [Medline: 27853384]
2. Franco Z, Fallaize R, Lovegrove JA, Hwang F. Popular Nutrition-Related Mobile Apps: A Feature Assessment. JMIR Mhealth Uhealth 2016 Aug 01;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
3. Waki K, Aizawa K, Kato S, Fujita H, Lee H, Kobayashi H, et al. DialBetics With a Multimedia Food Recording Tool, FoodLog: Smartphone-Based Self-Management for Type 2 Diabetes. J Diabetes Sci Technol 2015 May 16;9(3):534-540 [FREE Full text] [doi: 10.1177/1932296815579690] [Medline: 25883164]
4. Aizawa K, Ogawa M. FoodLog: Multimedia Tool for Healthcare Applications. IEEE MultiMedia 2015 Apr;22(2):4-8. [doi: 10.1109/MMUL.2015.39]
5. Yanai K, Kawano Y. Food image recognition using deep convolutional network with pre-training and fine-tuning. 2015 Jun 29 Presented at: IEEE International conference on Multimedia and Expo Workshops; ICMEW; 2015; Turin. [doi: 10.1109/icmew.2015.7169816]
6. Takeuchi T, Fujii T, Ogawa K, Narumi T, Tanikawa T, Hirose M. Using social media to change eating habits without conscious effort. 2014 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle p. 527-535. [doi: 10.1145/2638728.2641330]
7. The Japan Diabetes Society , editor. Food exchange Lists - Dietary Guidance for persons with Diabetes (in Japanese). Tokyo: Bunkodo; Nov 1, 2013.
8. Anthimopoulos J, Dehais J, Deim P, Mougiakakou S. Segmentation and recognition of multi food meal images for carbohydrate counting. 2013 Nov 10 Presented at: IEEE International conference on Bioinformatics and Bioengineering; BIBE; 2013; Chania. [doi: 10.1109/bibe.2013.6701608]
9. Yang S, Chen M, Pomerleau D, Sukhankar R. Food Recognition Using Statistics of Pairwise Local Features. 2010 Jun 13 Presented at: IEEE International conference on Computer Vision and Pattern Recognition; CVPR; 2010; San Francisco. [doi: 10.1109/cvpr.2010.5539907]
10. Farinella G, Naik N, Parikh D, Raskar R, Hidalgo CA. Classifying food images represented as bag of textons. 2014 Oct 27 Presented at: IEEE International conference on Image Processing; ICIP 2014; 2014; Paris. [doi: 10.1109/icip.2014.7026055]
11. Zhang W, Yu Q, Siddiquie B, Divakaran A, Sawhney H. "Snap-n-Eat": Food Recognition and Nutrition Estimation on a Smartphone. J Diabetes Sci Technol 2015 May;9(3):525-533 [FREE Full text] [doi: 10.1177/1932296815582222] [Medline: 25901024]
12. He H, Kong F, Tan J. DietCam: Multiview Food Recognition Using a Multikernel SVM. IEEE J Biomed Health Inform 2016 May;20(3):848-855. [doi: 10.1109/JBHI.2015.2419251] [Medline: 25850095]
13. Zhang X, Lu Y, Zhang S. Multi-Task Learning for Food Identification and Analysis with Deep Convolutional Neural Networks. J Comput Sci Technol 2016 May 6;31(3):489-500. [doi: 10.1007/s11390-016-1642-6]
14. Ming J, Chen Y, Cao C, Forde C, Ngo C, Tua T. Food photo recognition for dietary tracking; system and experiment. 2018 Feb 5 Presented at: 24th International conference on Multi Media Modeling; MMM; 2018; Bangkok.
15. He Y, Xu C, Khanna N. Food Image Analysis: Segmentation, identification and weight estimation. 2013 Presented at: IEEE International conference on Multimedia and Expo; ICME; July 15-19, 2013; San Jose. [doi: 10.1109/icme.2013.6607548]

16. Anthimopoulos MM, Gianola L, Scarnato L, Diem P, Mougiakakou SG. A food recognition system for diabetic patients based on an optimized bag-of-features model. IEEE J Biomed Health Inform 2014 Jul;18(4):1261-1271. [doi: 10.1109/JBHI.2014.2308928] [Medline: 25014934]
17. Chen M, Dhingra K, Wu W, Yang L, Sukthankar R, Yang J. PFID: Pittsburg fast-food image dataset. 2009 Presented at: IEEE International conference on Image Processing; ICIP 2009; November 7-10, 2009; Cairo. [doi: 10.1109/icip.2009.5413511]
18. Zhang X, Zhou F, Lin Y, Zhang S. Embedding label structures for fine-grained feature representation. 2016 Jun 27 Presented at: IEEE International conference on Computer Vision and Pattern Recognition; CVPR; 2016; Las Vegas. [doi: 10.1109/cvpr.2016.126]
19. Kawano Y, Yanai K. Foodcam 256: A large-scale real-time mobile food recognition system employing high dimensional features and compression of classifier weights. 2014 Presented at: ACM international Conference on Multimedia Modeling; ACM MM 2014; November 7, 2014; Orlando. [doi: 10.1145/2647868.2654869]
20. Ege T, Yanai K. Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions. 2017 Oct 23 Presented at: Thematic Workshops of ACM Multimedia; October 23-27, 2017; Mountain View p. 367-375. [doi: 10.1145/3126686.3126742]
21. Dubey A, Moltisanti M, Battiato S. Deep Learning the City: Quantifying Urban Perception at A Global Scale. 2016 Presented at: 14th European Conference on Computer Vision; ECCV; October 8-16, 2016; Amsterdam.
22. Herbrich R, Minka T, Graepel T. TrueSkill(TM): A Baysian Skill Rating System. 2006 Presented at: NIPS 2006; December 4-9, 2006; Vancouver.
23. Joachims T. Optimizing Search Engines using Clickthrough Data. Joachims; 2002 Presented at: ACM International conference on Knowledge Discovery and Data Mining; SIGKDD; July 23-26, 2002; San Francisco. [doi: 10.1145/775047.775067]
24. Hu MLY, Yu N. Multiple Instance Ranking: Learning to Rank Images for Image Retrieval. 2008 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; CVPR 2017; July 22-25, 2017; Honolulu. [doi: 10.1109/cvpr.2008.4587352]
25. Dietary Reference Intakes for Japanese (2015).: Health Service Bureau, Ministry of Health, Labour and Welfare URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-10900000-Kenkoukyoku/0000208954.pdf [accessed 2020-10-15]
26. Yoshiike N, Hayashi F, Takemi Y, Mizoguchi K, Seino F. A new food guide in Japan: the Japanese food guide Spinning Top. Nutr Rev 2007 Apr;65(4):149-154. [doi: 10.1111/j.1753-4887.2007.tb00294.x] [Medline: 17503709]
27. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. 2017 Jul 22 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; CVPR; July 22-25, 2017; Honolulu. [doi: 10.1109/cvpr.2017.660]

Abbreviations

CNN: convolutional neural network
MLRA: machine learning rank algorithm
PSPNET: pyramid scene parsing network
SVM: support vector machine

Edited by G Eysenbach; submitted 02.03.20; peer-reviewed by M Donaldson, S Capra; comments to author 01.06.20; revised version received 26.07.20; accepted 02.10.20; published 26.10.20.

Please cite as:
Sudo K, Murasaki K, Kinebuchi T, Kimura S, Waki K
Machine Learning–Based Screening of Healthy Meals From Image Analysis: System Development and Pilot Study
JMIR Form Res 2020;4(10):e18507
URL: http://formative.jmir.org/2020/10/e18507/
doi:10.2196/18507
PMID:33104010

Original Paper

Perceptions of Mobile Apps for Smoking Cessation Among Young People in Community Mental Health Care: Qualitative Study

Minda A Gowarty[1,2,3], MD; Nathan J Kung[2], BA; Ashley E Maher[2], BA; Meghan R Longacre[2,4], PhD; Mary F Brunette[2,3,5], MD

[1]Departments of Internal Medicine and Community and Family Medicine, Dartmouth Hitchcock Medical Center, Lebanon, NH, United States
[2]Geisel School of Medicine at Dartmouth, Hanover, NH, United States
[3]Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
[4]The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Hanover, NH, United States
[5]Department of Psychiatry, Dartmouth Hitchcock Medical Center, Lebanon, NH, United States

Corresponding Author:
Minda A Gowarty, MD
Departments of Internal Medicine and Community and Family Medicine
Dartmouth Hitchcock Medical Center
1 Medical Center Drive
Lebanon, NH, 03756
United States
Phone: 1 6036536868
Email: minda.a.gowarty@hitchcock.org

Abstract

Background: Young adults with serious mental illness are over twice as likely to have tobacco use disorder than those in the general population and are less likely to utilize proven treatment methods during quit attempts. However, little research has evaluated the efficacy of interventions for this group. Smartphone apps may be an underutilized tool for tobacco use disorder among young adults with serious mental illness.

Objective: The aim of this study was to explore attitudes toward smoking cessation apps and preferences regarding app design in young adult smokers with serious mental illness.

Methods: Five focus groups involving 25- to 35-year-old adults with serious mental illness receiving treatment at a community mental health center were conducted between May 2019 and August 2019. Three researchers independently coded transcripts and identified themes using thematic analysis.

Results: Participants (n=22) were individuals who smoke daily: 10 (46%) self-identified as female, 18 (82%) self-identified as White, and 9 (41%) had psychotic disorders. Key themes that emerged included a general interest in using health apps; a desire for apps to provide ongoing motivation during a quit attempt via social support, progress tracking, and rewards; a desire for apps to provide distraction from smoking; concerns about app effectiveness due to a lack of external accountability; and concerns that apps could trigger cravings or smoking behavior by mentioning cigarettes or the act of smoking.

Conclusions: Apps have the potential to support smoking cessation or reduction efforts among young adults with serious mental illness. However, they may require tailoring, optimization, and clinical support to effectively promote cessation in this population.

(JMIR Form Res 2020;4(10):e19860) doi:10.2196/19860

KEYWORDS

smoking cessation; mHealth; serious mental illness; smartphone application; digital health; psychiatric illness; tobacco treatment

Introduction

People with serious mental illness (disabling mood, anxiety, and psychotic disorders) are more likely to develop tobacco use disorders and are less likely to quit smoking than those in the general population, even when using recommended cessation interventions [1-3]. Tobacco smoking is a major contributor to high rates of chronic cardiovascular and lung diseases, high treatment costs (US $160 billion per year in the United States [4]), and to a 10- to 20-year reduction in life expectancy for people with serious mental illness [2,5-7]. While quitting smoking at any age confers health benefits, the harmful health

effects of smoking worsen as the number of cigarettes and duration of smoking increase [8]. Intervening at an early age can dramatically reduce the risk of smoking-related disease and can mitigate early mortality [8].

A significant body of literature [9-12] has described unique challenges faced by people with serious mental illness who wish to quit smoking, which include both sociocultural influences (such as lower education and income, as well as higher stress levels) and neurobiological influences (including the modulating effect of nicotine on dopaminergic pathways in schizophrenia). Additionally, people with serious mental illness may endorse using tobacco to manage challenging psychiatric symptoms and to overcome difficulties with socialization [10,12,13]. Furthermore, some cessation medications (eg, bupropion) may not be indicated for a subset of people with serious mental illness [14]. While overall, pharmacologic therapies are safe and effective for people with psychiatric illnesses [3,11,15-17], behavioral interventions are also needed to teach cessation skills, provide education about the safety and efficacy of smoking cessation medications, and support mental health. However, little research has focused on interventions for young adults with serious mental illness and tobacco use disorder, and effective strategies for addressing tobacco use disorder among young adults with serious mental illness have not yet been established.

Research on treatment in the general population is informative. Behavioral therapies for treating tobacco use disorder improve abstinence rates in youth and young adults in the general population, and are recommended by US Clinical Practice Guidelines [18]. A meta-analysis [19] that extracted data for young adult participants (18 to 24 years old) from studies of adult smokers found that behavioral interventions designed for older age groups were also effective for young adults. Research on interventions designed specifically for young adults found the most promising results from telephone and web-based interventions [20]. Yet evidence-based, scalable tobacco use disorder treatment approaches, such as telephone Quitline counseling, are underutilized by young people. For example, only 8.5% of US Quitline callers from 2011 to 2013 were young adults [21], and that percentage decreased to 5% in 2016 [22]. Young adults with and without serious mental illness report frequent quit attempts [23], but they are typically unaided and unsuccessful [24]. In our previous research [25], we found that young adults with serious mental illness, in contrast to middle-aged adults with serious mental illness, were less likely to initiate evidence-based tobacco use disorder treatment after education or standard motivational interviewing, suggesting that more appealing approaches to treatment are needed for this group.

Young people are avid users of smartphone technology [26], and smartphone apps offer beneficial features for behavioral intervention delivery. App content can be accessed on-demand, allowing users to capitalize on fleeting moments of motivation. Additionally, apps can deliver personalized and interactive content, including proactive notifications based on time and location [27]. In a recent survey [28], we found that 80% of young people with serious mental illness used smartphones, similar to the overall rate among young adults with low incomes (in the United States) [26]. Furthermore, 70% of young adults with serious mental illness were willing to try digital health interventions on their device [28], indicating preliminary feasibility for digital interventions for tobacco use disorder treatment in this group.

While smartphone apps offer a number of attributes that seem well-suited to young adults with and without serious mental illness, optimal development and implementation of this technology requires input from the intended users. However, phase I (design and refinement) and phase II (feasibility, proof-of-concept, or pilot testing) trials of smoking cessation apps are often omitted or are not reported in published literature [29]. Assessing end users' desires regarding app content and features is important in developing appealing interventions as well as for cost-effective implementation and reliable interpretation of effectiveness [29]. Recent work has addressed this gap in knowledge for middle aged adults with serious mental illness [30-33]. However, to our knowledge, research evaluating perceptions of apps for smoking cessation among young adults with serious mental illness has not been published.

Given the need for improved engagement in tobacco treatment among young adults with serious mental illness, and the promise of smartphone apps as an accessible and tailorable vehicle for behavioral intervention, we sought to explore attitudes among young adults with serious mental illness who smoke toward using apps for smoking cessation as well as preferences regarding app design. The goal of this study was to obtain information that could guide the tailoring of mobile app interventions to the unique needs of this population.

Methods

Participants and Recruitment

We used purposive sampling to recruit potentially eligible participants from a single large community mental health center in New England between May 2019 and August 2019. Recruitment occurred via flyers posted in waiting rooms and clinician referral. Eligible participants were 18 to 35 years old, English-speaking, stable in their outpatient mental health treatment for serious mental illness (ie, no hospitalization in past 30 days per chart review), and smoked daily. We chose the age range of 18 to 35 years to focus on adults for whom quitting smoking has the greatest potential long-term mortality benefit. While the National Young Adult Health Survey found that cessation attitudes were similar among 18- to 24-year-old and 25- to 34-year-old adults [34], it also demonstrated that smoking trajectories differ between these age groups [35]. Thus, we stratified patients by age into 2 groups—18 to 24 years and 25 to 35 years—to better characterize participants' cessation needs. This paper presents findings from the 25 to 35-year-old age group. Patients were excluded if they were pregnant or had a current, unstable substance use disorder (per chart review or per the participant's mental health center clinician).

Prior to each focus group, research staff read the study information sheet aloud with the participants and discussed the purpose of the study with them. All eligible participants were deemed competent to give consent, which participants provided

verbally. Participants received a US $30 gift card to a retail store after completing the focus group. The New Hampshire State Institutional Review Board approved and monitored all study activities.

Procedures

Brief Survey

Participants completed a 10-item survey prior to the start of focus group discussions. The survey included questions about the participants' gender identity, race, tobacco use, and technology use. Technology use included questions about app use, including whether participants had ever downloaded a health-related app (such as a step tracker or stress management app). Participants' DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) psychiatric diagnoses, as determined by their mental health center clinicians, were gathered by chart review.

Focus Groups

We conducted 5 focus groups between May 2019 and August 2019, each of which included 3 to 6 participants and lasted approximately one hour. A researcher trained in qualitative methods (MG) moderated the focus groups, with a second member of the research team present to take field notes. Focus groups took place at the community mental health center where participants received services.

The moderator posed questions in a funnel-type structure [36], starting with broad questions about participant experiences with smoking, narrowing to experiences with quitting or reducing smoking, and finally to perceptions about using smartphone apps to quit or reduce smoking. We followed a semistructured format, using probing and clarifying questions to elicit thick descriptions of participants' perceptions and experiences. The focus group discussions were audiorecorded and transcribed. A member of the research team who was present at the focus groups compared the transcriptions to the audio files to ensure accuracy. Focus groups were conducted until thematic saturation was reached, which was identified when previously recognized themes repeated without the emergence of new themes [37,38]. This occurred after the fifth focus group. The focus group discussion guide can be found in Multimedia Appendix 1.

Data Analysis

Transcripts were iteratively analyzed using thematic analytic techniques [39]. Three researchers (MG, NK, and AM) independently coded each transcript using ATLAS.ti (version 8, ATLAS.ti Scientific Software Development GmbH). After conducting an immersive review of the data set, an initial set of codes was generated using a deductive-inductive approach. This approach allowed us to generate codes based on prior empirical data on facilitators and barriers to quitting smoking, while also allowing new codes to emerge from the data set. The 3 researchers met regularly to refine the code definitions until reaching a final code structure, which they each independently applied to the entire data set. Additional meetings were held to discuss discrepancies in applying the final codes to the data set until consensus was reached through negotiation [40]. Memoing was used throughout the coding process to facilitate a deeper understanding of how the codes relate to each other, and how codes could be represented by unifying themes [41]. The themes were then developed into thematic statements, and emblematic quotations where chosen to illustrate how the themes developed from the data. Negative case analysis was used to ensure the entire data set was represented in the emerging themes.

Results

Study Participants

Participants (n=22) were individuals who smoked daily and who were stable in their community mental health treatment. Almost half of participants (10/22, 46%) self-identified as female, most (18/22, 82%) self-identified as white, and 9/22 (41%) were diagnosed with psychotic disorders. Technology use data are presented in Table 1.

Table 1. Focus group participant characteristics.

Characteristic	Value, n (%)
Demographic and clinical characteristics (n=22)	
Gender	
Male	12 (54)
Female	10 (46)
Race	
White	18 (82)
Mixed	3 (14)
I don't know	1 (5)
Psychotic disorder diagnosis	9 (41)
Technology use characteristics (n=19[a])	
Use internet ≥ twice daily	18 (95)
Ever downloaded an app	18 (95)
Ever downloaded a health app	16 (84)
Would try an app if recommended by a doctor	15 (79)

[a]Technology use data were missing for 3 participants.

Focus Group Themes

Facilitators and Barriers to Quitting Smoking

During the initial discussions about reasons for smoking and participants' prior experiences with quitting, a number of themes emerged regarding facilitators and barriers to quitting. Reasons for quitting included the desire to save money and concern regarding their children's, pets', and personal health. For example, one participant stated:

> *I feel like whenever I quit smoking, I can smell things easier, I can taste things more, and I can breathe better.* [Group 1 Participant 3]

Commonly cited reasons to continue smoking despite a desire to quit included addiction to nicotine, smoking as routine, and smoking to manage mental health symptoms (such as stress, anxiety, and depression):

> *...since I have bipolar, it helps. Smoking helps me not have so much anxiety, but when I do quit, I get even more.* [Group 5 Participant 1]

Relapse triggers included socializing with other smokers, smelling tobacco smoke, and seeing people smoke on television or in movies. Of note, while many participants mentioned prior use of nicotine replacement therapy or cessation medications such as varenicline or bupropion with varying degrees of success, most stated that either they or their clinicians were not comfortable with using prescription cessation medications due to concern for psychiatric side effects. For example:

> *I took Chantix for a while...it really did help. I was smoking like two packs a day, and then when I was on the Chantix I was smoking maybe six cigarettes a day...and a doctor told me that it might contribute to depression. And I decided to get off it because I already have depression, and...I don't want to worsen it.* [Group 4 Participant 1]

> *That's what I was told. None of my doctors, nobody will give me Chantix. Because of how bad I am.* [Group 4 Participant 2]

Role of Apps in Supporting a Quit or Reduction Attempt

Focus group discussions regarding apps included participants' prior experiences with health-related apps as well as their perceptions regarding the role of smoking cessation apps. Their prior experiences were generally positive, though with variable effect on behavior. As participants discussed mobile apps for smoking cessation, many expressed interest in using apps during a quit attempt and described a number of ways that apps could offer support, such as providing motivation during a quit attempt, increasing awareness of smoking habits and money spent, and providing distraction from smoking. They also described app limitations, such as lack of external accountability and the potential to trigger cravings. These themes, along with illustrative quotations, are summarized in Table 2. Subthemes that emerged from these discussions are subsequenty presented.

Table 2. Major themes regarding app use for smoking cessation, with corresponding illustrative quotations.

Themes	Illustrative quotation
Suggested ways apps could support a quit attempt	
Providing ongoing motivation	"I think it'd be cool if you can be able to, like, challenge someone else who was trying to quit smoking, but like a buddy, right? But it's through the app. And they could be thousands of miles away, but you got that one person..." [Group 5 Participant 2]
Increasing awareness of smoking patterns	"Tracking when you smoke, the times you smoke, how much a day you smoke, what led you to smoke that much. Like, I feel like those are all helpful things to know" [Group 4 Participant 1]
Providing distraction	"Yeah, 'cause when you're having a craving, you just look at it [the app] and maybe it'll tell you, like, uh, go for an hour run, or you know, tell you some sort of structure to keep your mind off of you smoking. Something to keep you busy, keep your hands busy..." [Group 3 Participant 1]
App limitations in supporting a quit attempt	
Lack of external accountability	"...how is this going to know when I'm smoking a cigarette or not? I can just say I'm not...and then I'll be sitting there smoking a cig, you know" [Group 2 Participant 4]
Possible triggers	"You'd have to use, like a code word for cigarette so people don't think it in their heads because once they think 'cigarette,' they're more likely to smoke" [Group 2 Participant 3]

Prior Experiences With Apps

Many participants described interest in health-related apps such as fitness trackers or mood trackers but noted varying degrees of benefit during their prior experiences with these types of apps. Participants discussed using the apps to review information about their personal habits (ie, their "stats"), but altering a behavior based on this feedback was rare. For example, a participant using a step tracker said,

> *It was cool, a cool thing. It didn't really matter because I was working every day and I'd get about the same amount of steps every day.* [Group 1 Participant 4]

Only one participant described increasing her activity level to achieve goals in her step tracker; others simply reviewed their steps without this leading to change in behavior. Another participant found a symptom tracker useful to bring to her clinicians so they could change her treatment plan but did not utilize the information herself.

Motivation

Participants felt that while motivation to quit is a prerequisite for using a smoking cessation app, the app would need to provide features that could facilitate ongoing motivation. Participants suggested a number of ways that apps could motivate them, such as receiving support from other people within the app, feedback about progress (such as cigarettes avoided or money saved), and rewards such as financial incentives or badges. They also mentioned the importance of distraction to avoid cigarettes and suggested that an app could offer distraction by providing suggestions for alternate activities instead of smoking or by including games within the app.

Social Support

While one participant was concerned that an app would lead to missed social interactions and reduced social support for quitting, others discussed desire for the app to incorporate social support. Multiple participants suggested the ability to connect with other smokers for support within the app. One person suggested a chat feature:

> *So I think if you could, like, message someone on the app that's using it at the same time...if you could communicate with someone else using it, like chat with them...* [Group 3 Participant 5]

Other participants recommended group challenges, similar to those seen in popular fitness apps:

> *I think it'd be cool if you can be able to, like, challenge someone else who was trying to quit smoking, but like a buddy, right? But it's through the app. And they could be thousands of miles away, but you got that one person...* [Group 5 Participant 2]

While most of the participants who valued social support in their quit attempt mentioned other smokers or peers, one suggested involving family members as part of the quit plan in the app:

> *Or someone, a family member or a loved one, could work with the app and you could earn points or something like that by doing that more than smoking.* [Group 2 Participant 3]

Tracking Progress

A prominent theme across focus groups was the desire to track the number of cigarettes smoked per day. Many participants wanted to track cigarettes to increase their awareness of how much they smoke. Others added that they would like to record additional information about their smoking so they could learn their typical smoking patterns:

> *Tracking when you smoke, the times you smoke, how much a day you smoke, what led you to smoke that*

> *much. Like, I feel like those are all helpful things to know.* [Group 4 Participant 1]

Some participants noted that quitting all at once can feel insurmountable, but suggested that tracking cigarettes could be motivating by allowing them to see progress toward quitting:

> *Maybe not to quit, because to me, quitting is not realistic. If I could reduce, like, instead of smoking 20 cigarettes, if I could smoke 5 cigarettes a day, that would've been a big difference.* [Group 2 Participant 1]

Most participants cited money as a major motivator to quit or reduce their smoking and would want the app to provide a feature that tracks money spent on cigarettes or money saved by avoiding cigarettes:

> *I should, I feel like what'd probably help me, if like, if I kept track to see how much I'm paying, spending on them [cigarettes] because... if I kept track to see how much I'm paying every time, I'd be like, 'Okay, that needs to stop.' Because that's money that I could be using.* [Group 4 Participant 3]

Skills Training

While tracking was a common theme, fewer participants suggested skill-based features for behavioral change. One participant recommended that an app include

> *Something to disassociate the triggers and patterns associated with smoking.* [Group 1 Participant 5]

Another stated that

> *The app has to be informative. It should have tips on how to reduce the urge to smoke.* [Group 2 Participant 1]

However, such statements were far less common than recommendations for tracking.

Rewards

A common theme across focus groups was the desire for rewards within the app. Some participants mentioned financial rewards such as gift cards for using the apps while others recommended that the app award badges for progress, based on their experience with other health behavior change apps:

> *I think that those badges, you know, when I get those rewards...nobody else sees them. I'm the only one, but it is a reward to myself...It's just mine. Nobody can take it from me. This is stuff that I worked hard for, and worked, probably, pretty hard for.* [Group 5 Participant 2]

Distraction

Another prominently mentioned feature across focus groups was the benefit of distraction during a quit attempt. Many participants described using distraction as a tool for avoiding cigarettes during prior quit attempts and felt that a smartphone app would be well-suited to providing distraction. Some participants suggested that the app could provide games to play to avoid smoking. Others felt that the app could provide suggestions for alternative activities to smoking:

> *Yeah, 'cause when you're having a craving, you just look at it [the app] and maybe it'll tell you, like, uh, go for an hour run, or you know, tell you some sort of structure to keep your mind off of you smoking. Something to keep you busy, keep your hands busy...* [Group 3 Participant 1]

App Limitations

Most participants felt that the biggest barrier to using an app to quit or reduce smoking was a lack of motivation to change smoking behavior in general, and that once motivated, there were few barriers other than practical issues related to any mobile technology (eg, limited phone battery, inability to use phone if lost or broken, limited cellular or wireless internet service in certain locations). However, two main concerns arose regarding the limitations of apps during a quit attempt—a lack of external accountability and the potential to trigger cravings.

A prominent theme across focus groups was the need for external accountability during a quit attempt and concern that an app would not be able to provide this. Related to their discussion of the importance of monitoring progress, participants voiced concerns about the temptation to report false information to an app. One participant, who had previously used a cigarette tracking app, noted:

> *I tried a cigarette counter once, and realized it wasn't gonna work to help me quit smoking because I, at that time, I would cheat and not log all of my cigarettes. I don't even know why I cheated because it's not like anyone was watching.* [Group 1 Participant 2]

In another focus group, a similar sentiment arose:

> *...how is this going to know when I'm smoking a cigarette or not? I can just say I'm not...and then I'll be sitting there smoking a cig, you know.* [Group 2 Participant 4]

Other participants felt that apps might have limited benefit without a method of confirming smoking status. One participant stated:

> *Like, the phone has to have a contact system where you're, like, Face-Timing somebody who's coming to visit you to go meet for coffee and then they're smelling you and they can see if you smoked any cigarettes by doing that test...* [Group 2 Participant 3]

Despite general agreement that entering information into the app on an honor system is a limitation, participants who mentioned biochemical verification as a means to achieve this did so unfavorably. In one group, a participant described mistrust of the accuracy of breath carbon monoxide monitoring. In another focus group, a participant who mentioned concern about false reporting to the app went on to consider breath carbon monoxide monitoring, but then immediately discounted it, saying,

> *...that's a little much...* [Group 2 Participant 4]

Participants made few recommendations about features that should be avoided in a smoking cessation app, but a common

concern was the possibility that the app could trigger cravings. Based on their prior experiences with smoking triggers, they were concerned that any mention of cigarettes in the app could increase their desire to smoke:

> *You'd have to use, like a code word for cigarette so people don't think it in their heads because once they think 'cigarette,' they're more likely to smoke.* [Group 2 Participant 3]

Discussion

Principal Results and Comparison With Prior Work

In this study, we explored the attitudes of young adults with serious mental illness who smoke toward quitting smoking and the use of apps for this purpose. Overall, our findings share significant overlap with those of prior early phase trials in middle-aged adults with serious mental illness [30,33]. We found that young adults with serious mental illness who smoke shared similar reasons for quitting or continuing to smoke as those in other populations and were interested in using apps during a quit or reduction attempt. Participants voiced a desire for apps to provide ongoing motivational content during a quit attempt, features to increase awareness of smoking habits and money spent, content that could be used as a distraction from smoking, reward features, and social support for quitting smoking. They also described potential app limitations, such as the temptation to enter false information into the tracking features, and an app's potential to trigger cravings. This group's concern about psychiatric side effects of cessation pharmacotherapy is an additional important characteristic that could be countered or monitored with digital technology.

Our analysis revealed a tension between participants' desire to see progress via tracking features and their fear of seeing personal failure via the same features. Similar to qualitative findings in studies of middle-aged adults with and without mental illness [30,33,42-44], participants in our study voiced a strong desire for cigarette- and money-tracking features that could demonstrate progress and enhance motivation during a quit attempt. However, participants also described using other apps with similar features that had not affected their patterns of behavior. Additionally, and similar to other adults with and without mental health issues [33,45], they expressed concern that recording information that suggests a lack or loss of progress might be demotivating and could lead to relapse or to the temptation to enter false information into the app. Although tracking smoking and viewing progress were popular features noted by this study group and those in other research [30,42,43,46], in one study of general population adult daily smokers [43], these features were not associated with improved abstinence.

Similar to other studies of adults with serious mental illness, in general [10,12,13,33], and young people with mental health conditions, in particular [47], participants described their main barriers to quitting as inability to resist cravings, using smoking to manage mental health symptoms, and smoking out of habit. Yet, only a few suggested that apps could teach strategies to overcome these barriers. Instead, most participants conveyed the common perception that it just takes "will power" to quit, rather than the application of skills to tolerate stress and urges without smoking. Skills training, such as advice on changing routines and improving coping strategies for cravings, has been shown to improve abstinence outcomes for both in-person [48] and app-based [43] interventions. Taken together, these data suggest that tracking features are appealing and may enhance engagement or motivation, but other skill building or clinical support features for cessation skills would need to be prominent and engaging to ensure use within the apps.

Many participants in this study recommended that cessation apps include financial rewards or badges awarded by the app for progress. There is a growing body of literature to support the use of rewards in the form of praise [48] and financial incentives [49,50] for smoking cessation. Inclusion of rewards features in smoking cessation apps, therefore, has promise to improve both engagement and abstinence outcomes.

While prior studies assessing users' desire for specific features in smoking cessation apps demonstrate mixed results regarding the role of social support for quitting within apps [33,42,44,45], participants in our study clearly valued this feature. In line with prior findings regarding the role of social influence on smoking behavior in people with serious mental illness [10,51], our participants highlighted the importance of the effects of social environment and peer interaction on their smoking behaviors and described the potential benefit of an alternative smoke-free support system within the app. They specifically voiced a desire for peer support through chat functions and competitions, as opposed to smoking cessation coaching or technology coaching that has been described in other studies [30,32]. This desire for peer support is similar to preferences described by young adults with serious mental illness in a study [52] assessing the possible role of digital support for mental health diagnoses and highlights the perceived importance of peer-to-peer interaction for overcoming the stigma and challenges associated with smoking as well as living with mental illness. Although one study [53] demonstrated the potential of a motivational intervention delivered by peers, Dickerson and colleagues [54] described numerous challenges to cessation associated with a peer mentoring approach for people with serious mental illness. While this feature may improve engagement in the app and promote overall well-being, its effect on smoking behaviors requires further study.

The landscape of smoking cessation apps is rapidly changing. Currently available apps are variable in their content and features [27,29,55-58], and few contain content that adheres to clinical practice guidelines [27,55-57,59-61]. A recent literature review [29] found that most smoking cessation apps included self-tracking features, but only one-third included social support and one-third included rewards systems. Furthermore, most apps included three or fewer features, which typically involved education, tracking, and a variable third feature [29]. Another recent review [58] found that less than half of available apps included advice on changing routines or coping with cravings. Apps that include high-quality, evidence-based content as well as an array of both desired features and features previously found to be effective should be tested in this population.

Compared to other populations of smokers, the young adults with serious mental illness in this study voiced similar reasons for smoking, quitting smoking, and relapsing after a quit attempt [34,62,63], but they expressed specific concerns regarding medication safety in relation to their mental health conditions. A wide body of evidence has demonstrated the safety and efficacy of cessation pharmacotherapy in the setting of mental illness [3,11,15-17]. Thus, improving utilization of these treatments can substantially impact quit rates. Our findings suggest that content in a smoking cessation app tailored for young adults with serious mental illness can likely overlap substantially with content in other smoking cessation apps, with additional information about medication safety for people with mental illness.

Limitations

Several study limitations merit consideration. First, participants were not required to be interested in quitting smoking in order to participate in the study; individuals who are actively engaged in an attempt to quit smoking may ultimately prefer different app features. Second, participants were predominantly White residents in a small New England city, and may not be representative of smokers with serious mental illness who have other demographic characteristics or are from other geographic regions. Yet the substantial overlap between our findings and those of research in other populations [30,33,42,43] supports the validity and generalizability of the themes conveyed here. Lastly, most participants had not previously used apps for smoking cessation, and therefore, relied on anticipated future desires for their responses. While anticipation of future desires can be subject to a number of biases, assessing participants' preferences prior to exposure to specific apps is important to understanding how they will appraise apps during initial use. Future work should assess responses to cessation apps in this population.

Conclusions

Young adults with serious mental illness expressed similar reasons for quitting smoking or continuing to smoke compared to those expressed by other populations of individuals who smoke, but low cessation treatment utilization rates and quit rates suggest that other treatment modalities are needed. Apps have the potential to support quit or reduction attempts for young adults with serious mental illness in a number of ways, such as providing ongoing motivation during a quit attempt, increasing awareness of smoking habits and money spent, and providing information and support for using cessation skills. However, while young adult with serious mental illness who smoke are interested in using apps, further tailoring, optimization, and clinical support may be necessary to effectively promote cessation in this population.

Acknowledgments

The authors would like to thank Susan Guarino and Jason Welsh for their assistance with recruiting participants. MAG's work is supported by National Institute on Drug Abuse (Grant Number P30DA029926) via a pilot award through the Center for Technology and Behavioral Health at Dartmouth, and by a Health Resources Services Administration postdoctoral fellowship grant (T32HP32520). MFB's work is supported by Substance Abuse and Mental Health Services Administration (Grant Number 1H79SM080245-01). These funding organizations had no involvement in the study design; the collection, analysis, or interpretation of the data; the preparation of the report; or the decision to submit the article for publication.

Authors' Contributions

MAG and MFB designed the study with qualitative expertise provided by MRL. MAG, NJK, and AEM collected data and performed data analysis, with oversight of these activities provided by MFB and MRL. MAG and MFB prepared the original draft of the manuscript. MFB, NJK, AEM, and MRL reviewed and provided comments on the manuscript before submission.

Conflicts of Interest

MFB has received research funding support from Alkermes. The other authors have no competing interests to declare.

Multimedia Appendix 1
Focus group discussion guide.
[DOCX File , 16 KB - formative_v4i10e19860_app1.docx]

References

1. Smith PH, Mazure CM, McKee SA. Smoking and mental illness in the U.S. population. Tob Control 2014 Nov;23(e2):e147-e153 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051466] [Medline: 24727731]
2. Wang TW, Asman K, Gentzke AS, Cullen KA, Holder-Hayes E, Reyes-Guzman C, et al. Tobacco product use among adults - United States, 2017. MMWR Morb Mortal Wkly Rep 2018 Nov 09;67(44):1225-1232 [FREE Full text] [doi: 10.15585/mmwr.mm6744a2] [Medline: 30408019]
3. Anthenelli RM, Benowitz NL, West R, St Aubin L, McRae T, Lawrence D, et al. Neuropsychiatric safety and efficacy of varenicline, bupropion, and nicotine patch in smokers with and without psychiatric disorders (EAGLES): a double-blind,

randomised, placebo-controlled clinical trial. The Lancet 2016 Jun;387(10037):2507-2520. [doi: 10.1016/s0140-6736(16)30272-0]
4. Xu X, Bishop EE, Kennedy SM, Simpson SA, Pechacek TF. Annual healthcare spending attributable to cigarette smoking: an update. Am J Prev Med 2015 Mar;48(3):326-333 [FREE Full text] [doi: 10.1016/j.amepre.2014.10.012] [Medline: 25498551]
5. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry 2015 Apr;72(4):334-341 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.2502] [Medline: 25671328]
6. Chesney E, Goodwin G, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. World Psychiatry 2014 Jun;13(2):153-160 [FREE Full text] [doi: 10.1002/wps.20128] [Medline: 24890068]
7. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med 2013 Jan 24;368(4):341-350. [doi: 10.1056/NEJMsa1211128] [Medline: 23343063]
8. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 2004 Jun 26;328(7455):1519 [FREE Full text] [doi: 10.1136/bmj.38142.554479.AE] [Medline: 15213107]
9. Ziedonis D, Hitsman B, Beckham JC, Zvolensky M, Adler LE, Audrain-McGovern J, et al. Tobacco use and cessation in psychiatric disorders: National Institute of Mental Health report. Nicotine Tob Res 2008 Dec;10(12):1691-1715. [doi: 10.1080/14622200802443569] [Medline: 19023823]
10. Tidey JW, Miller ME. Smoking cessation and reduction in people with chronic mental illness. BMJ 2015 Sep 21;351:h4065 [FREE Full text] [doi: 10.1136/bmj.h4065] [Medline: 26391240]
11. Cather C, Pachas GN, Cieslak KM, Evins AE. Achieving smoking cessation in individuals with schizophrenia: special considerations. CNS Drugs 2017 Jun;31(6):471-481 [FREE Full text] [doi: 10.1007/s40263-017-0438-8] [Medline: 28550660]
12. Lum A, Skelton E, Wynne O, Bonevski B. A systematic review of psychosocial barriers and facilitators to smoking cessation in people living with schizophrenia. Front Psychiatry 2018;9:565 [FREE Full text] [doi: 10.3389/fpsyt.2018.00565] [Medline: 30459658]
13. Trainor K, Leavey G. Barriers and facilitators to smoking cessation among people with severe mental illness: a critical appraisal of qualitative studies. Nicotine Tob Res 2017 Jan;19(1):14-23. [doi: 10.1093/ntr/ntw183] [Medline: 27613905]
14. Joffe RT, MacQueen GM, Marriott M, Robb J, Begin H, Young LT. Induction of mania and cycle acceleration in bipolar disorder: effect of different classes of antidepressant. Acta Psychiatr Scand 2002 Jun;105(6):427-430. [doi: 10.1034/j.1600-0447.2002.02360.x] [Medline: 12059846]
15. Tsoi DT, Porwal M, Webster AC. Interventions for smoking cessation and reduction in individuals with schizophrenia. Cochrane Database Syst Rev 2013;2:CD007253. [doi: 10.1002/14651858.CD007253.pub3] [Medline: 23450574]
16. Evins AE, Cather C, Pratt SA, Pachas GN, Hoeppner SS, Goff DC, et al. Maintenance treatment with varenicline for smoking cessation in patients with schizophrenia and bipolar disorder: a randomized clinical trial. JAMA 2014 Jan 08;311(2):145-154 [FREE Full text] [doi: 10.1001/jama.2013.285113] [Medline: 24399553]
17. George TP, Vessicchio JC, Sacco KA, Weinberger AH, Dudas MM, Allen TM, et al. A placebo-controlled trial of bupropion combined with nicotine patch for smoking cessation in schizophrenia. Biol Psychiatry 2008 Jun 01;63(11):1092-1096 [FREE Full text] [doi: 10.1016/j.biopsych.2007.11.002] [Medline: 18096137]
18. Fiore M, Jaén C, Baker T. Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: U.S. Department of Health and Human Services; May 2008.
19. Suls JM, Luger TM, Curry SJ, Mermelstein RJ, Sporer AK, An LC. Efficacy of smoking-cessation interventions for young adults: a meta-analysis. Am J Prev Med 2012 Jun;42(6):655-662 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.013] [Medline: 22608385]
20. Villanti AC, McKay HS, Abrams DB, Holtgrave DR, Bowie JV. Smoking-cessation interventions for U.S. young adults: a systematic review. Am J Prev Med 2010 Dec;39(6):564-574. [doi: 10.1016/j.amepre.2010.08.009] [Medline: 21084078]
21. Marshall LL, Zhang L, Malarcher AM, Mann NH, King BA, Alexander RL. Race/ethnic variations in Quitline use among US adult tobacco users in 45 states, 2011-2013. Nicotine Tob Res 2017 Nov 07;19(12):1473-1481. [doi: 10.1093/ntr/ntw281] [Medline: 29121347]
22. National quitline data warehouse statistical brief. Smoking & Tobacco Use.: Centers for Disease Control and Prevention; 2018 Apr 25. URL: https://www.cdc.gov/tobacco/quit_smoking/cessation/nqdw/statistical-brief/index.htm [accessed 2020-01-28] [WebCite Cache ID https://www.cdc.gov/tobacco/quit_smoking/cessation/nqdw/statistical-brief/index.htm]
23. Brunette MF, Feiron JC, Aschbrenner K, Colctti D, Devitt T, Greene MA, et al. Characteristics and predictors of intention to use cessation treatment among smokers with schizophrenia: young adults compared to older adults. J Subst Abus Alcohol 2017;5(1) [FREE Full text] [Medline: 29881770]
24. Brunette MF, Ferron JC, Aschbrenner KA, Pratt SI, Geiger P, Kosydar S. Attitudes about smoking cessation treatment, intention to quit, and cessation treatment utilization among young adult smokers with severe mental illnesses. Addict Behav 2019 Feb;89:248-255 [FREE Full text] [doi: 10.1016/j.addbeh.2018.09.028] [Medline: 30343187]

25. Brunette MF, Ferron JC, Robinson D, Coletti D, Geiger P, Devitt T, et al. Brief web-based interventions for young adult smokers with severe mental illnesses: a randomized, controlled pilot study. Nicotine Tob Res 2018 Sep 04;20(10):1206-1214. [doi: 10.1093/ntr/ntx190] [Medline: 29059417]
26. Mobile Fact Sheet. Pew Research Center. Washington D.C: Pew Research Center; 2018 Feb 05. URL: https://www.pewinternet.org/fact-sheet/mobile/
27. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, et al. How smart are smartphone apps for smoking cessation? a content analysis. Nicotine Tob Res 2015 Jun 4. [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
28. Brunette M, Achtyes E, Pratt S, Stilwell K, Opperman M, Guarino S, et al. Use of smartphones, computers and social media among people with smi: opportunity for intervention. Community Ment Health J 2019 Jun 8;55(6):973-978. [doi: 10.1007/s10597-019-00431-7]
29. Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile applications for the treatment of tobacco use and dependence. Curr Addict Rep 2019 May 9;6(2):86-97. [doi: 10.1007/s40429-019-00248-0]
30. Vilardaga R, Rizo J, Kientz JA, McDonell MG, Ries RK, Sobel K. User experience evaluation of a smoking cessation app in people with serious mental illness. Nicotine Tob Res 2016 May;18(5):1032-1038. [doi: 10.1093/ntr/ntv256] [Medline: 26581430]
31. Vilardaga R, Rizo J, Ries RK, Kientz JA, Ziedonis DM, Hernandez K, et al. Formative, multimethod case studies of learn to quit, an acceptance and commitment therapy smoking cessation app designed for people with serious mental illness. Transl Behav Med 2019 Nov 25;9(6):1076-1086 [FREE Full text] [doi: 10.1093/tbm/iby097] [Medline: 30445507]
32. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
33. Klein P, Lawn S, Tsourtos G, van Agteren J. Tailoring of a smartphone smoking cessation app (kick.it) for serious mental illness populations: qualitative study. JMIR Hum Factors 2019 Sep 03;6(3):e14023 [FREE Full text] [doi: 10.2196/14023] [Medline: 31482850]
34. Villanti AC, Bover Manderski MT, Gundersen DA, Steinberg MB, Delnevo CD. Reasons to quit and barriers to quitting smoking in US young adults. Fam Pract 2016 Apr;33(2):133-139 [FREE Full text] [doi: 10.1093/fampra/cmv103] [Medline: 26733658]
35. Delnevo CD, Villanti AC, Wackowski OA, Gundersen DA, Giovenco DP. The influence of menthol, e-cigarettes and other tobacco products on young adults' self-reported changes in past year smoking. Tob Control 2016 Sep;25(5):571-574 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052325] [Medline: 26243809]
36. Morgan D. Focus Groups as Qualitative Research Second Edition. Thousand Oaks, CA: SAGE Publications, Inc; 1997.
37. Morse JM. The significance of saturation. Qualitative Health Research 1995 May 01;5(2):147-149. [doi: 10.1177/104973239500500201]
38. Guest G, Namey E, McKenna K. How many focus groups are enough? building an evidence base for nonprobability sample sizes. Field Methods 2016 Jul 24;29(1):3-22. [doi: 10.1177/1525822X16639015]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
41. Padgett D. Qualitative and Mixed Methods in Public Health. Thousand Oaks, CA: SAGE Publications, Inc; 2012.
42. Oliver JA, Hallyburton MB, Pacek LR, Mitchell JT, Vilardaga R, Fuemmeler BF, et al. What do smokers want in a smartphone-based cessation application? Nicotine Tob Res 2017 Aug 03. [doi: 10.1093/ntr/ntx171] [Medline: 29065202]
43. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015 Jan;41(1):68-73. [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
44. Paay J, Kjeldskov J, Skov MB, Lichon L, Rasmussen S. Understanding individual differences for tailored smoking cessation apps. New York, NY: Association for Computing Machinery; 2015 Presented at: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April 2015; Seoul Republic of Korea p. 1699-1708. [doi: 10.1145/2702123.2702321]
45. Thornton LK, Kay-Lambkin FJ. Specific features of current and emerging mobile health apps: user views among people with and without mental health problems. mHealth 2018 Dec;4:56-56. [doi: 10.21037/mhealth.2018.11.04]
46. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
47. Prochaska JJ, Fromont SC, Wa C, Matlow R, Ramo DE, Hall SM. Tobacco use and its treatment among young people in mental health settings: a qualitative analysis. Nicotine Tob Res 2013 Aug;15(8):1427-1435 [FREE Full text] [doi: 10.1093/ntr/nts343] [Medline: 23322765]

48. West R, Walia A, Hyder N, Shahab L, Michie S. Behavior change techniques used by the English Stop Smoking Services and their associations with short-term quit outcomes. Nicotine Tob Res 2010 Jul;12(7):742-747. [doi: 10.1093/ntr/ntq074] [Medline: 20478957]
49. Notley C, Gentry S, Livingstone-Banks J, Bauld L, Perera R, Hartmann-Boyce J. Incentives for smoking cessation. Cochrane Database of Systematic Reviews 2019;7. [doi: 10.1002/14651858.cd004307.pub6]
50. Brunette MF, Pratt SI, Bartels SJ, Scherer EA, Sigmon SC, Ferron JC, et al. Randomized trial of interventions for smoking cessation among Medicaid beneficiaries with mental illness. Psychiatr Serv 2018 Mar 01;69(3):274-280. [doi: 10.1176/appi.ps.201700245] [Medline: 29137560]
51. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. J Am Coll Cardiol 2019 Jun 25;73(24):e285-e350. [doi: 10.1016/j.jacc.2018.11.003] [Medline: 30423393]
52. Aschbrenner KA, Naslund JA, Tomlinson EF, Kinney A, Pratt SI, Brunette MF. Adolescents' use of digital technologies and preferences for mobile health coaching in public mental health settings. Front Public Health 2019;7:178 [FREE Full text] [doi: 10.3389/fpubh.2019.00178] [Medline: 31312629]
53. Williams JM, Dwyer M, Verna M, Zimmermann MH, Gandhi KK, Galazyn M, et al. Evaluation of the CHOICES program of peer-to-peer tobacco education and advocacy. Community Ment Health J 2011 Jun;47(3):243-251. [doi: 10.1007/s10597-010-9310-8] [Medline: 20419349]
54. Dickerson F, Savage CLG, Schweinfurth LAB, Goldberg RW, Bennett M, Dixon L, et al. The experience of peer mentors in an intervention to promote smoking cessation in persons with psychiatric illness. Community Ment Health J 2016 May;52(4):416-423 [FREE Full text] [doi: 10.1007/s10597-015-9967-0] [Medline: 26602772]
55. Ferron JC, Brunette MF, Geiger P, Marsch LA, Adachi-Mejia AM, Bartels SJ. Mobile phone apps for smoking cessation: quality and usability among smokers with psychosis. JMIR Hum Factors 2017 Mar 03;4(1):e7 [FREE Full text] [doi: 10.2196/humanfactors.5933] [Medline: 28258047]
56. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
57. Robinson CD, Seaman EL, Grenen E, Montgomery L, Yockey RA, Coa K, et al. A content analysis of smartphone apps for adolescent smoking cessation. Transl Behav Med 2020 Feb 03;10(1):302-309. [doi: 10.1093/tbm/iby113] [Medline: 30476293]
58. Ubhi HK, Michie S, Kotz D, van SOCP, Selladurai A, West R. Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features. Transl Behav Med 2016 Sep;6(3):410-417 [FREE Full text] [doi: 10.1007/s13142-015-0352-x] [Medline: 27528530]
59. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
60. Abroms LC, Lee WJ, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
61. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
62. Twyman L, Bonevski B, Paul C, Bryant J. Perceived barriers to smoking cessation in selected vulnerable groups: a systematic review of the qualitative and quantitative literature. BMJ Open 2014 Dec 22;4(12):e006414 [FREE Full text] [doi: 10.1136/bmjopen-2014-006414] [Medline: 25534212]
63. McCaul KD, Hockemeyer JR, Johnson RJ, Zetocha K, Quinlan K, Glasgow RE. Motivation to quit using cigarettes: a review. Addict Behav 2006 Jan;31(1):42-56. [doi: 10.1016/j.addbeh.2005.04.004] [Medline: 15916861]

Edited by G Eysenbach; submitted 05.05.20; peer-reviewed by E Grenen, S Lawn; comments to author 16.07.20; revised version received 14.08.20; accepted 17.08.20; published 02.10.20.

Please cite as:
Gowarty MA, Kung NJ, Maher AE, Longacre MR, Brunette MF
Perceptions of Mobile Apps for Smoking Cessation Among Young People in Community Mental Health Care: Qualitative Study
JMIR Form Res 2020;4(10):e19860
URL: https://formative.jmir.org/2020/10/e19860
doi:10.2196/19860
PMID:33006560

Original Paper

Expanding Access to Perinatal Depression Treatment in Kenya Through Automated Psychological Support: Development and Usability Study

Eric P Green[1]; Yihuan Lai[1]; Nicholas Pearson[1,2]; Sathyanath Rajasekharan[3]; Michiel Rauws[4]; Angela Joerin[4]; Edith Kwobah[5]; Christine Musyimi[6]; Rachel M Jones[3]; Chaya Bhat[1]; Antonia Mulinge[3]; Eve S Puffer[1,7]

[1]Duke Global Health Institute, Durham, NC, United States
[2]Jacaranda Health, San Francisco, CA, United States
[3]Jacaranda Health, Nairobi, Kenya
[4]X2AI, San Francisco, CA, United States
[5]Moi Teaching and Referral Hospital, Eldoret, Kenya
[6]Africa Mental Health Research and Training Foundation, Nairobi, Kenya
[7]Department of Psychology and Neuroscience, Duke University, Durham, NC, United States

Corresponding Author:
Eric P Green
Duke Global Health Institute
Box 90519
Durham, NC, 27708
United States
Phone: 1 9196817289
Email: eric.green@duke.edu

Abstract

Background: Depression during pregnancy and in the postpartum period is associated with poor outcomes for women and their children. Although effective interventions exist for common mental disorders that occur during pregnancy and the postpartum period, most cases in low- and middle-income countries go untreated because of a lack of trained professionals. Task-sharing models such as the *Thinking Healthy* Program have shown potential in feasibility and efficacy trials as a strategy for expanding access to treatment in low-resource settings; however, there are significant barriers to scale-up. We address this gap by adapting *Thinking Healthy* for automated delivery via a mobile phone. This new intervention, *Healthy Moms*, uses an existing artificial intelligence system called Tess (Zuri in Kenya) to drive conversations with users.

Objective: This prepilot study aims to gather preliminary data on the *Healthy Moms* perinatal depression intervention to learn how to build and test a more robust service.

Methods: We conducted a single-case experimental design with pregnant women and new mothers recruited from public hospitals outside of Nairobi, Kenya. We invited these women to complete a brief, automated screening delivered via text messages to determine their eligibility. Enrolled participants were randomized to a 1- or 2-week baseline period and then invited to begin using Zuri. We prompted participants to rate their mood via SMS text messaging every 3 days during the baseline and intervention periods, and we used these preliminary repeated measures data to fit a linear mixed-effects model of response to treatment. We also reviewed system logs and conducted in-depth interviews with participants to study engagement with the intervention, feasibility, and acceptability.

Results: We invited 647 women to learn more about Zuri: 86 completed our automated SMS screening and 41 enrolled in the study. Most of the enrolled women submitted at least 3 mood ratings (31/41, 76%) and sent at least 1 message to Zuri (27/41, 66%). A third of the sample engaged beyond registration (14/41, 34%). On average, women who engaged post registration started 3.4 (SD 3.2) *Healthy Moms* sessions and completed 3.1 (SD 2.9) of the sessions they started. Most interviewees who tried Zuri reported having a positive attitude toward the service and expressed trust in Zuri. They also attributed positive life changes to the intervention. We estimated that using this alpha version of Zuri may have led to a 7% improvement in mood.

Conclusions: Zuri is feasible to deliver via SMS and was acceptable to this sample of pregnant women and new mothers. The results of this prepilot study will serve as a baseline for future studies in terms of recruitment, data collection, and outcomes.

International Registered Report Identifier (IRRID): RR2-10.2196/11800

(JMIR Form Res 2020;4(10):e17895) doi:10.2196/17895

KEYWORDS

telemedicine; mental health; depression; artificial intelligence; Kenya; text messaging; mobile phone

Introduction

Depression is a leading cause of disability worldwide. Women experiencing perinatal depression are a particularly underserved population. Depression during pregnancy and in the postpartum period (perinatal depression) affects as many as 20% of women in high-income countries [1] and may be more prevalent in low- and middle-income countries (LMICs) [2]. The condition is associated with a number of poor outcomes in women and their children, including increased maternal morbidity and mortality [3,4], poor infant health [5-9], and poor developmental outcomes [10-12].

Although effective interventions exist for common mental disorders that occur during pregnancy and the postpartum period [13], most cases in LMICs go untreated. In these settings, more than 7 of 10 people who need treatment cannot access care because of a lack of trained professionals [14]. In Kenya, for example, there are only 180 psychiatric nurses outside of the capital city, a ratio of about 1 provider per 200,000 to 250,000 people. To close this gap, the World Health Organization developed the Mental Health Gap Action Programme (mhGAP) intervention guide, which outlines how to deliver mental health services in primary health care settings through nonspecialist providers. This task-sharing approach has proved efficacious, particularly for maternal mental health [15].

One example of an intervention based on the mhGAP intervention guide is the 15-session *Thinking Healthy Program*, a cognitive behavioral therapy (CBT)–based intervention for treating perinatal depression that is intentionally nonstigmatizing (eg, uses words such as *stress* and *burden* instead of *depression* and *illness*) [16]. Community health workers—typically women educated through secondary school with no specific background in mental health—are trained over 5 to 10 days to help pregnant women learn three skills: to identify unhealthy thinking, to replace unhealthy thinking with helpful thinking, and to practice thinking and acting healthy. In a trial in Pakistan with 900 pregnant women, Rahman et al [17] found that the intervention halved the prevalence of major depression, and a 7-year follow-up study reported a persistent effect of treatment (as well as some spontaneous recovery among the control group) [18]. A peer-delivered version of *Thinking Healthy* may offer an alternative, cost-effective strategy for treating perinatal depression [19].

Despite this impressive evidence of feasibility and efficacy, there are significant barriers to scale-up [20], and there is evidence that the effects of *Thinking Healthy* might not extend to children of depressed mothers without additional engagement [21]. Common implementation challenges of task-sharing models include a lack of funding and infrastructure for training and service delivery, workforce retention in the absence of compensation or incentives for nonspecialists, high workloads, transportation costs, appointment scheduling logistics, and inadequate clinical supervision [22]. Although it is critical to study how to optimize and scale these task-sharing approaches, the fact remains that today, most women in LMICs who need treatment still have no access to care.

Given this treatment gap and barriers to scale-up, our intention is to make it possible for anyone with a basic mobile phone (ie, a feature phone with only text messaging capabilities) to receive high-quality, evidence-based psychological support anytime, anywhere. We are attempting this in the context of perinatal depression by adapting *Thinking Healthy* to an existing artificial intelligence (AI) system for automated psychological support called Tess, which we have named *Zuri* in Kenya. This idea is innovative because it introduces an entirely new delivery channel that has the potential for a step change in expanding access to care while also potentially augmenting and strengthening existing task-sharing models.

Zuri works by engaging a patient in conversation via a variety of trusted channels, including text messaging (SMS). Either Zuri or the patient can start a conversation, and Zuri can be programmed to walk a patient through a structured curriculum such as *Thinking Healthy*. As a safety measure, conversations with patients in need of additional support can be handed over to live counselors as needed. The potential benefits of this approach include on-demand 24/7 access for an unlimited number of patients, no scheduling of appointments, no travel costs to appointments, enhanced sense of privacy and avoidance of social stigma, and high fidelity to treatment.

Our long-term goal is to expand access to high-quality, on-demand treatment services to people who have common mental disorders such as perinatal depression but cannot receive care from mental health professionals because of cost and human resource constraints. The main objectives of this study are to adapt *Thinking Healthy* for dissemination in Kenya through the Zuri AI system, develop and test study procedures to inform the design of a randomized controlled trial (RCT), and generate preliminary evidence of feasibility, acceptability, and response to treatment.

Methods

Research Design

We adapted *Thinking Healthy* for the Zuri AI system and evaluated the combined perinatal depression intervention, which we are calling *Healthy Moms*, with a cohort of pregnant women and new mothers recruited from 2 large public hospitals in Kenya. We used a single-case experimental design (partially nonconcurrent multiple baseline [23], open label) and qualitative interviews to generate preliminary data on feasibility, acceptability, and response to treatment. This is a stage 2

Registered Report. The stage 1 protocol (DERR1-10.2196/11800) describes our preliminary work to adapt *Thinking Healthy* for dissemination in Kenya through the Zuri AI system [24].

Participants and Recruitment

We recruited pregnant women and new mothers from 2 large public hospitals in Kiambu County, Kenya (population approximately 2.5 million, 60% urban). Both hospitals are part of a county-wide partnership offering patients innovative SMS programs that promote healthy motherhood [25]. When a woman signed up for the county SMS, we sent her an invitation via SMS to complete an automated SMS screening (in English) to determine if she was eligible for *Healthy Moms*. The screening included questions about age, maternity status, expected or actual delivery date, 9 questions about symptoms of depression from the Patient Health Questionnaire–9 (PHQ-9) [26], and a question about her current mood.

We informed all women who completed the automated screening that a study team member would call them within 1 business day. During this follow-up call, women who endorsed having thoughts of self-harm in the previous 2 weeks (question 9 on the PHQ-9) were offered a referral for counseling but were not eligible to enroll in *Healthy Moms*, given the early stage of intervention development. All other women were eligible to enroll as long as they confirmed that they were at least 20 weeks pregnant or no more than 6 months postpartum. The study coordinator (AM)—a Kenyan woman fluent in English and Swahili—assessed each woman's English-speaking ability on the call and asked women to rate their ability to read and understand English. Women could enroll regardless of language ability; however, we informed women with low English literacy that they might not find value in the current version of the program if they were not comfortable reading and writing in English.

If a woman chose to continue the enrollment process, the study coordinator read the informed consent form, answered her questions, and obtained verbal informed consent to enroll. The study coordinator asked enrollees to share information about the type of phone they use, schooling, number of dependents, marital status, and employment status.

Eligibility

To be eligible to participate, women needed to meet the following criteria: (1) pregnant (>20 weeks) or less than 6 months postpartum, (2) receiving antenatal or postnatal health care services from a participating hospital in Kiambu County, (3) have access to any type of mobile phone, (4) be enrolled in the county SMS program, and (5) be at least 18 years of age. English language proficiency and self-reported experience of depression symptoms were not required but were assessed. Women who endorsed suicidal ideation at the time of recruitment were ineligible to enroll in the study and were informed about potential resources for treatment.

Randomization to Baseline Length

As each woman enrolled in the study, we attempted to match her to another new enrollee of similar maternity status and randomly assigned the pair to have a 1-week or 2-week baseline period (using a random number generator). The intention was to ensure that every participant had a concurrent baseline period with at least one other person.

Intervention

We developed the *Health Moms* intervention based on the original *Thinking Healthy* manual for community health workers [16]. We also created a companion *Healthy Moms* journal that we printed and delivered to enrolled participants [27]. The journal included modified health calendars from the original *Thinking Healthy* manual along with short session summaries and writing prompts. This prepilot study was an opportunity to get feedback on the journal to ascertain how we might adapt the content into text, audio, and video for electronic delivery (and ultimately discontinue print versions in future trials). We conducted an initial round of user testing to develop the SMS intervention journal content [28].

Unlike *Thinking Healthy*, which trains community health workers to deliver the in-person intervention to women in need, we designed *Healthy Moms* for automated delivery via text messaging. We maintained the *Thinking Healthy* structure of 15 sessions overall: 3 prenatal sessions and 12 postnatal sessions during the first 10 months of the infant's life.

When it was time for a woman to participate in a *Healthy Moms* session, we (Zuri) sent her a text message to let her know that a new session was ready. The automated session began when she replied via SMS (Later in the study, we also enabled women to chat with Zuri via Facebook Messenger.). Each automated session followed the same 4-task format as that of *Thinking Healthy*: (1) reviewing key lessons from the previous session, (2) reviewing her mood ratings, (3) teaching new skills, and (4) introducing practice-based homework. Multimedia Appendices 1 and 2 provide an example *Healthy Moms* session conversation flow and associated journal content.

In between *Healthy Moms* sessions, women were encouraged to start a conversation with Zuri by asking a question or saying "Hi." Zuri attempted to discern the user's request and responded automatically with answers or replies that used active listening techniques such as restatement and reflection.

During this *free chat* mode, Zuri would ask a question similar to "How are you feeling now?" If the response indicated neutral or positive emotions, Zuri would offer a rapport-building module (eg, music, cooking, passions). If the response indicated a negative emotion, Zuri would offer a supportive intervention (eg, mindfulness and relaxation). Module selection was prioritized on the basis of aggregate helpfulness ratings from all user interactions in the X2AI/Tess system so that the most helpful modules were offered first. There was no limit to how much or how often a user could engage with Zuri.

If a woman discussed self-harm or other crisis topics, Zuri alerted a live study support member who could take over the chat session or call the participant directly and facilitate a referral to traditional in-person treatment if indicated (Zuri was programmed to inform women that her response might not be immediate at this stage of testing; therefore, they should seek help at an emergency room if in a crisis.). During enrollment,

we also informed participants that they were free to seek concomitant care and interventions at any point during the study.

Just as mental health specialists and nonspecialists trained to deliver psychotherapy improve over time with practice and experience, AI-enhanced systems such as Zuri also change, albeit in more subtle ways, given the current state of the technology. For instance, Zuri's emotion recognition algorithms updated automatically when it correctly or incorrectly interpreted the emotional valence of a user's input; however, the didactic intervention content did not change dynamically. Modifications to the intervention content were made manually; we reviewed conversation transcripts and made minor changes to the wording or sequence of messages when we noticed that users were confused or not engaging.

Outcomes and Data Collection Procedures

We collected data on study implementation, intervention engagement, feasibility and acceptability of the intervention, and patient outcomes, including depression severity and current mood.

Study Implementation

We tracked data on the recruitment funnel from the initial screening invite through the secondary eligibility screening to ultimate engagement with the intervention. We also tracked participants' responses to regular prompts to complete automated assessments throughout the study period.

Intervention Engagement

We assessed intervention engagement by reviewing Zuri system logs to document the completion of *Healthy Moms* sessions and patient-initiated engagement with Zuri outside of scheduled sessions. The Zuri system logs also informed our assessment of feasibility and acceptability; low engagement was considered a marker of potential barriers to feasibility or a lack of acceptability.

Feasibility and Acceptability of the Intervention

We further explored feasibility and acceptability by inviting 15 enrolled women to participate in individual interviews during the evaluation period. We purposively invited 3 different types of participants: those who did not finish the registration process with Zuri (n=5), those who finished the registration process but did not complete a session (n=5), and those who completed at least one session (n=5). A master's level trainee (YL) and the study coordinator (AM, Kenyan) conducted the interviews. Women who did not complete a full session with Zuri were interviewed over telephone. Women who completed one or more sessions were reimbursed to travel to one of the study hospitals for an in-person interview. The interviews lasted approximately 20 to 40 min and were based on a semistructured interview guide. The guide included open-ended questions and follow-up probes related to reasons for using Zuri, attitudes toward Zuri, favorite features, preferences of language and platform, challenges encountered, and perceived impact after using Zuri. The interviews were conducted in English; however, the study coordinator provided simultaneous translation to Swahili as needed.

In addition to these interviews, we also attempted to document all contacts the research team had with participants outside of the Zuri AI system and logged all adverse events. We were interested in determining how much assistance or encouragement users need from the team to understand and use the automated intervention.

Patient Outcomes

To measure mood, we asked participants to rate their feelings on a 10-point scale that we created and tested with users [24], where 1 meant very sad and 10 meant very happy (shifted to 0-9 for analysis). We invited women to rate their current mood via SMS during the enrollment screening and then every 3 days throughout the baseline and intervention periods. Each rating invitation reminded women of their previous rating. We also encouraged women to track and reflect on their mood and behaviors on a daily basis using the *Healthy Moms* journal we provided as part of the intervention (not analyzed) [27].

We also administered the PHQ-9 [26] via SMS. Our intention was to assess depression severity throughout the intervention period. However, after developing the protocol, we determined that the depression screening was too long to administer on a repeating basis. Instead, we opted to collect our minimum target of 2 self-ratings of depression severity, representing pre- and posttreatment.

Empirical Approach

Describe Study Implementation and Intervention Engagement

We used the study database to summarize the recruitment funnel and outcome data collection progress. We quantified intervention engagement in several ways. First, we used the system logs to summarize how frequently each participant engaged with the intervention by either participating in a *Healthy Moms* session (in response to a scheduled invite) or initiating a chat with Zuri in between the scheduled sessions. We also calculated and summarized the delay between our invitations to begin a *Healthy Moms* session and participants' start times, the proportion of *Healthy Moms* sessions started and completed, and the duration of participant-initiated chats with Zuri.

Explore Intervention Feasibility and Acceptability

As a hypothesis-generating exercise, we estimated the magnitude and direction of the associations between participant characteristics measured at baseline (eg, age, education, literacy, and symptom severity) and intervention engagement by fitting a Bayesian linear regression model.

We also explored barriers to and facilitators of engagement during in-depth interviews with participants and reviews of chat transcripts. Throughout the process, the analyst (YL) wrote memos to capture the main themes. In preparation for the thematic analysis, she developed a codebook and randomly selected one transcript that was double-coded and discussed. After refining the codebook, she used NVivo 12 (QSR International) to code memos and transcripts. The analyst wrote analytic memos for each thematic code, identifying similarities and differences across transcripts using a constant comparative

method [29]. She identified representative quotations of each theme.

Generate Preliminary Evidence About Participants' Response to Treatment

We aggregated the individual N-of-1 studies and estimated the magnitude of response and quantified uncertainty by fitting Bayesian linear mixed-effects models [30] in R (version 3.5) using the brms package [31] with default priors. As described in the protocol, the first model we fit included a random effect for observations nested within participants and the following fixed effects: (1) an intercept, (2) a dummy indicator for the treatment phase, (3) a time-within-baseline variable centered around the first observation (equal to 0 for observations outside of the baseline period), and (4) a time-within-treatment variable centered around the last observation (equal to 0 for observations outside of the treatment period). We applied a first-order autoregressive structure on the covariance matrix for the within-person residuals to account for autocorrelation.

We also fit a similar model not described in the protocol that reflected a lesson we learned in another project: rather than centering the time-within-period variables around a single observation, it may be more reasonable to center around the average of several consecutive observations when there is substantial individual variability in daily ratings. In this model, we centered the time-within-baseline variable around the *first 3* observations and centered the time-within-treatment variable around the *last 3* observations. Given the data availability, this 3-observation centering window was practical; we did not run the model with different window sizes to avoid cherry-picking the results. In the end, our choice of centering had no impact on the results; therefore, we decided to focus on the 3-observation centering window as an example of what we would likely attempt in a future trial using this design.

We augmented this quantitative analysis with a qualitative analysis of the in-depth interviews. We explored what links, if any, participants could make between engagement with the intervention and their mood, health, and relationships. We also intended to explore themes among women who did not exhibit positive changes in mood (*nonresponders*); however, this was not feasible, given the delay in launching this study.

Research Ethics

We obtained approvals to conduct this study from the institutional review boards at Duke University (US, 2018-0396) and Strathmore University (Kenya, SU-IRB 0210/18) and from the National Commission for Science and Technology in Kenya.

The study coordinator, AM (female, Kenyan, bachelor's degree), explained the study to prospective participants via telephone, administered the informed consent procedures, and obtained women's oral consent to enroll in this study.

Study participants were provided with an honorarium of up to Ksh 1500 (roughly US $15) delivered via mobile money transfer to recognize the time spent in completing study assessments. The original plan was to make these transfers after women completed sessions 1, 5, and 10; however, in practice, we sent women prorated honoraria on the basis of lower benchmarks of engagement, given the delay in launching this study.

X2AI, the creators of the AI system that we used to deliver *Healthy Moms*, transferred data to the research team in accordance with X2AI's data security policies [32]. The first author (EG) stored identifiable study data on a secure server during the study and then deidentified the data for analysis using the Safe Harbor method. Anonymized quantitative data and the code used to generate this manuscript are available for reanalysis [33].

Summary of Deviations From Stage 1 Protocol

In addition to changing the tense of the writing from future to past, we also made several edits to the *Introduction* section and modified several procedures described in the *Methods* section of the stage 1 protocol [24]: we (1) labeled the study as *prepilot* rather than *pilot* to better reflect that the data are preliminary and intended to inform the design of a larger pilot study; (2) moved text from the *Scientific Objectives and Significance* and *Expected Outcomes* sections to the *Discussion* section (but did not alter the objectives); (3) expanded access to the intervention from just SMS to include Facebook Messenger; (4) visualized the daily mood ratings but relied on model fitting rather than visual inspection to estimate trends and period impacts; and (5) dropped a planned *nonresponder* qualitative inquiry and modified the honorarium schedule because of limited time.

Results

Recruitment and Participants

We invited 647 women (446/647, 68.9% pregnant; 201/647, 31.1% new mothers) already enrolled in their county's SMS program to learn more about Zuri; 13.3% (86/647) of women completed our automated SMS screening between February 12, 2019, and June 18, 2019 (15/86, 17% of all women screened scored at or above the cutoff for possible depression; mean 9.5, SD 4.9). We determined that 52 of these 86 women were eligible to participate; 41 of 52 women completed the enrollment process (Figure 1).

Figure 1. Study flow diagram.

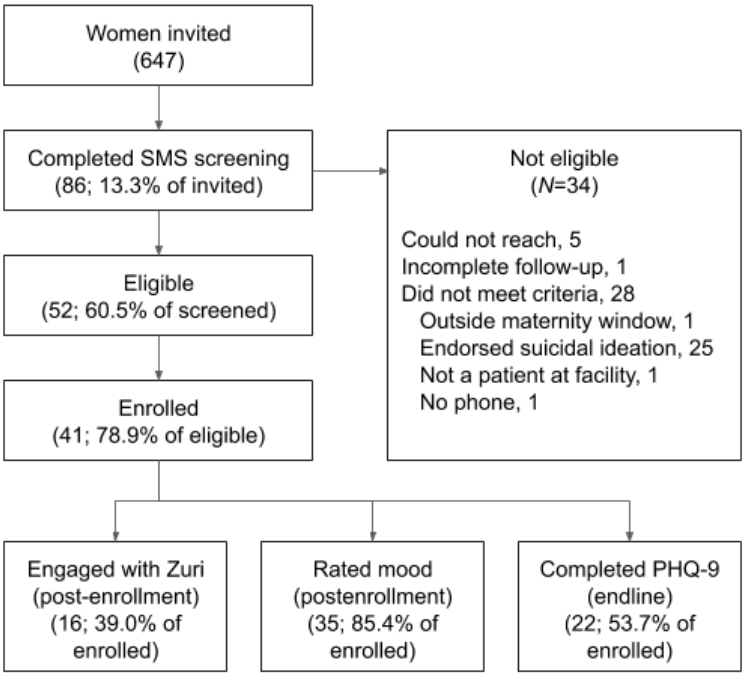


Table 1 reports the characteristics of the enrolled participants. The sample was evenly divided between pregnant women and new mothers. The average woman enrolled in the study was 25.9 years old (SD 4.8). All women reported that they could read in English, and the study interviewer reported that all could speak English. Most women used a smartphone, attended secondary school or higher, were married, and did not work. Women were not recruited on the basis of depression symptoms, and only 1 had a PHQ-9 score ≥15 at the time of enrollment [34]. The average PHQ-9 score upon study entry was 8.2 (possible maximum value of 27), and the average mood rating was 7.8 (possible maximum value of 9).

We conducted interviews with 15 of the 41 women enrolled in the study. They ranged in age from 20 to 38 years. Most were married and had delivered their baby within the last 6 months. All of the interviewees attended some secondary schooling, and 2 had earned a bachelor's degree.

Table 1. Characteristics of participants.

Characteristics	All women (n=41)	Maternity status: pregnant (19/41, 46%)	Maternity status: postpartum (22/41, 54%)
Age (years), mean (SD)	25.9 (4.8)	24.3 (3.1)	27.2 (5.5)
Self-reported English language reading skills, n (%)			
Poor	0 (0)	0 (0)	0 (0)
Just okay	0 (0)	0 (0)	0 (0)
Good	12 (30)	7 (37)	5 (23)
Excellent	28 (68)	12 (63)	16 (73)
Missing	1 (2)	0 (0)	1 (2)
Highest level of school attended, n (%)			
None	0 (0)	0 (0)	0 (0)
Primary	0 (0)	0 (0)	0 (0)
Postprimary or vocational	0 (0)	0 (0)	0 (0)
Secondary	22 (54)	14 (74)	8 (36)
College	11 (27)	3 (16)	8 (36)
University	7 (17)	2 (10.5)	5 (23)
Missing	1 (2)	0 (0)	1 (4.5)
Phone type: smartphone, n (%)	33 (81)	14 (74)	19 (86)
Employed outside the home: no, n (%)	32 (78)	15 (79)	17 (77)
Number of dependent children, mean (SD)	1.1 (0.9)	0.5 (0.5)	1.6 (0.9)
Marital status, n (%)			
Single	3 (7)	1 (5)	2 (9)
Separated	0 (0)	0 (0)	0 (0)
Cohabiting	0 (0)	0 (0)	0 (0)
Married	37 (90)	18 (95)	19 (86)
Missing	1 (2)	0 (0)	1 (4.5)
PHQ-9[a] total score, possible 0-27, mean (SD)	8.2 (3.6)	8.7 (4.1)	7.8 (3.2)
Possible depression: (PHQ-9≥15), n (%)	1 (2)	1 (5)	0 (0)
Mean mood at enrollment, possible 0-9, mean (SD)	6.8 (2.4)	7.1 (2.4)	6.6 (2.4)

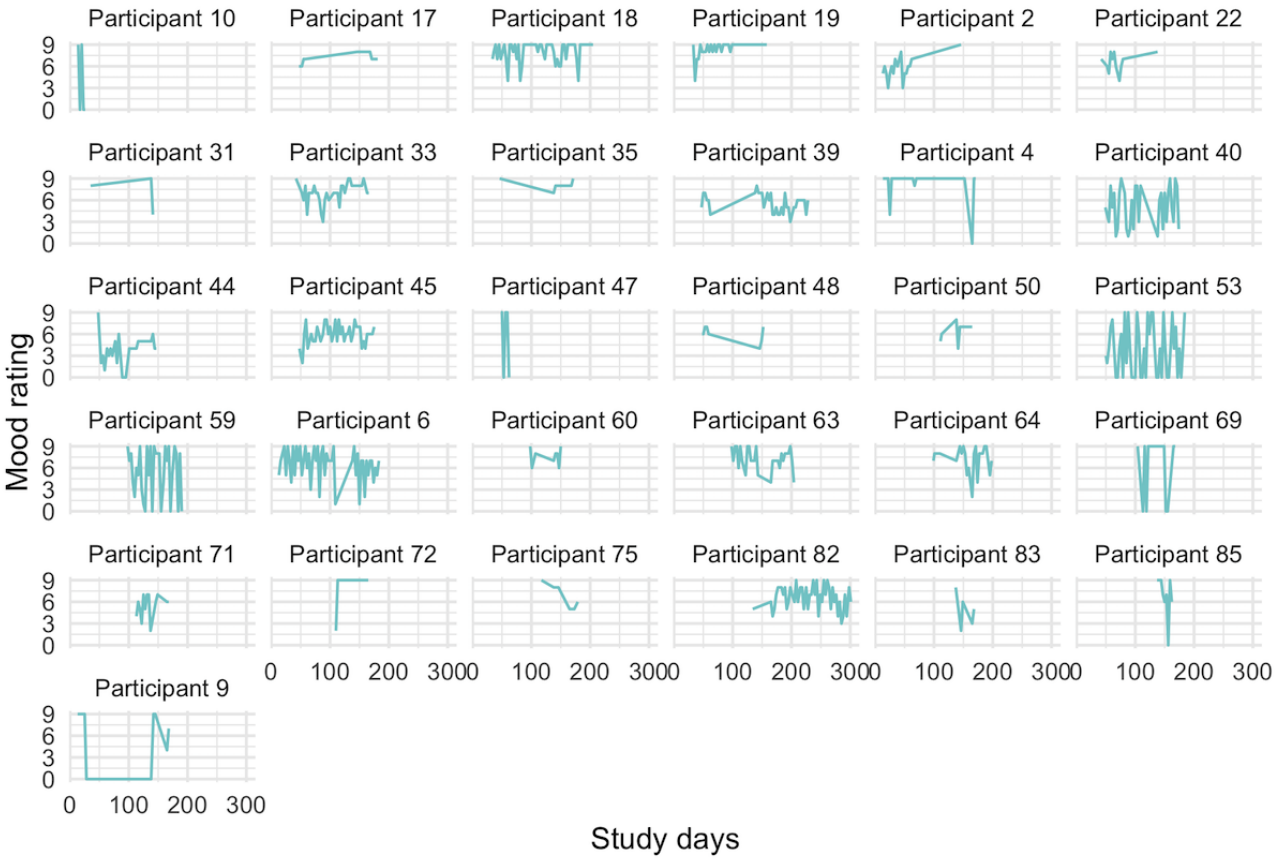
[a]PHQ-9: Patient Health Questionnaire–9.

Data Collection

Mood Ratings

Overall, the enrolled women submitted 719 daily mood ratings over the course of the study. The average woman submitted 17.5 ratings (SD 17.2), and 76% (31/41) of women submitted at least 3 ratings. The grand mean mood rating was 6.4 of 9 (SD 1.3) among those who submitted at least 3 ratings. Figure 2 suggests that most women reported a high degree of variability in ratings from one day to the next.

Figure 2. Time series of 705 mood ratings among 31 participants who submitted at least three ratings.



PHQ-9

We did not attempt to administer the PHQ-9 on a regular, ongoing basis to avoid frustrating users and distracting from potential engagement with the intervention. Instead, we only requested that women complete the PHQ-9 again at the end of the study period; 54% (22/41) of women responded.

Intervention Feasibility and Acceptability

Engagement Patterns

Over the course of the study, 66% (27/41) women sent at least one message to Zuri to begin the registration process, and 34% (14/41) of these women engaged with the intervention content beyond registration. Among this postregistration engagement subset, the average woman engaged with Zuri on 7.7 days (SD 6.0) and sent 130.5 messages (SD 117.4). On average, women sent 36.4% of these messages to Zuri in free chat mode, not as part of a *Healthy Moms* session. The median conversation unfolded over 0.6 hours (range 0.0-14.6 hours). Figure 3 displays the distributions of these engagement metrics.

To further investigate the nature of participant-initiated chats, we analyzed conversation transcripts and summarized the conversation modules engaged. Figure 4 shows the distribution of incoming messages by the free chat conversation module and maternity status. The most common rapport building module asked users about their passion in life. The most common intervention module outside of the *Healthy Moms* content was mindfulness-based meditation. In general, pregnant women were more likely to engage in intervention content during free chats compared with new mothers. This means that after rapport-building chats, Zuri suggested an intervention module and the women agreed to try.

On average, women who engaged in Zuri postregistration started 3.4 (SD 3.2) *Healthy Moms* sessions and completed 3.1 (SD 2.9) of the sessions they started. The median time from a *push* session invite to a woman responding was 0.6 hours (range 0.0-740.1 hours). Figure 5 shows a woman's engagement pattern over the course of the study period. There were no reported adverse events. One woman's conversation was flagged in real time for a potential crisis follow-up.

Figure 3. Distribution of number of days engaged and number of incoming messages sent among 14 women who engaged with Zuri beyond registration.

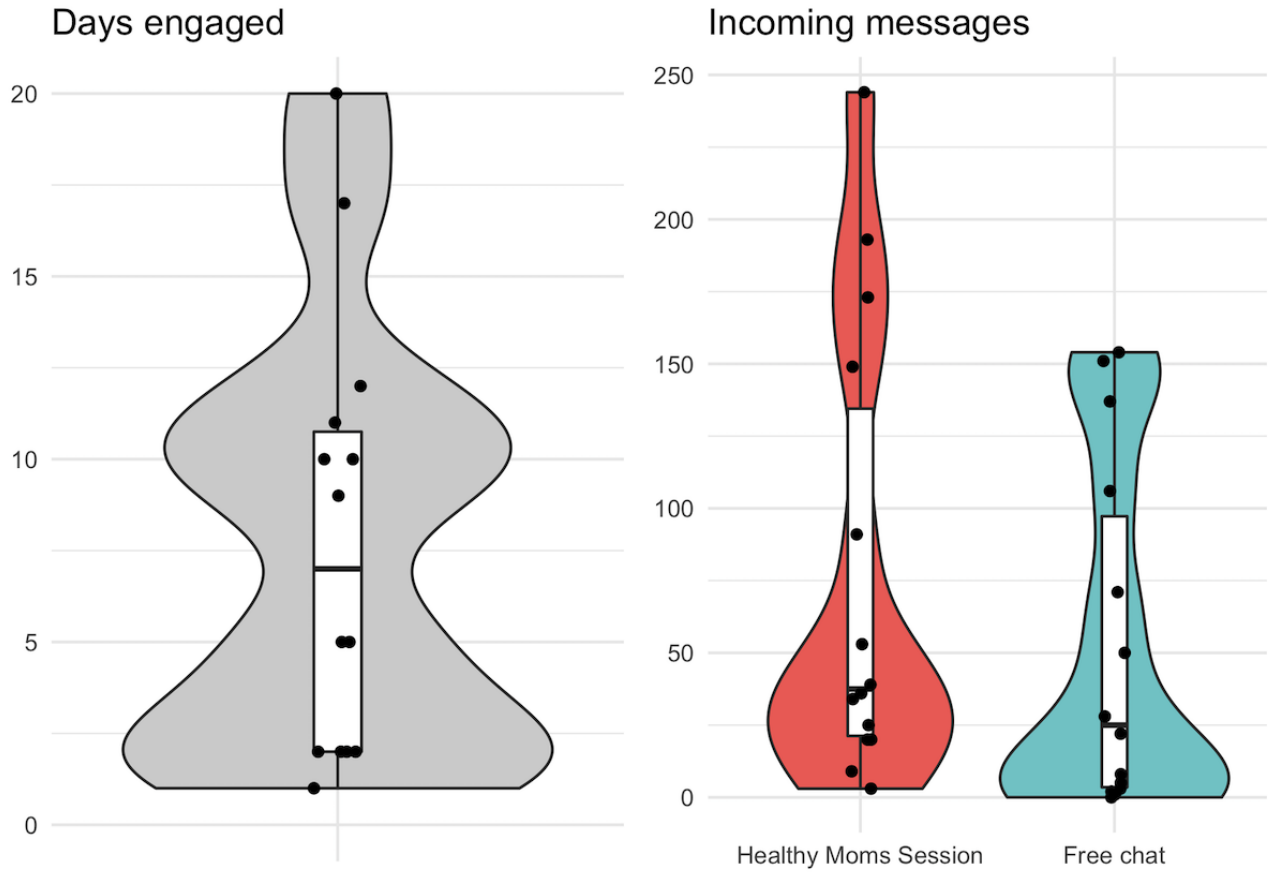


Figure 4. Distribution of incoming messages by free chat conversation module and maternity status.

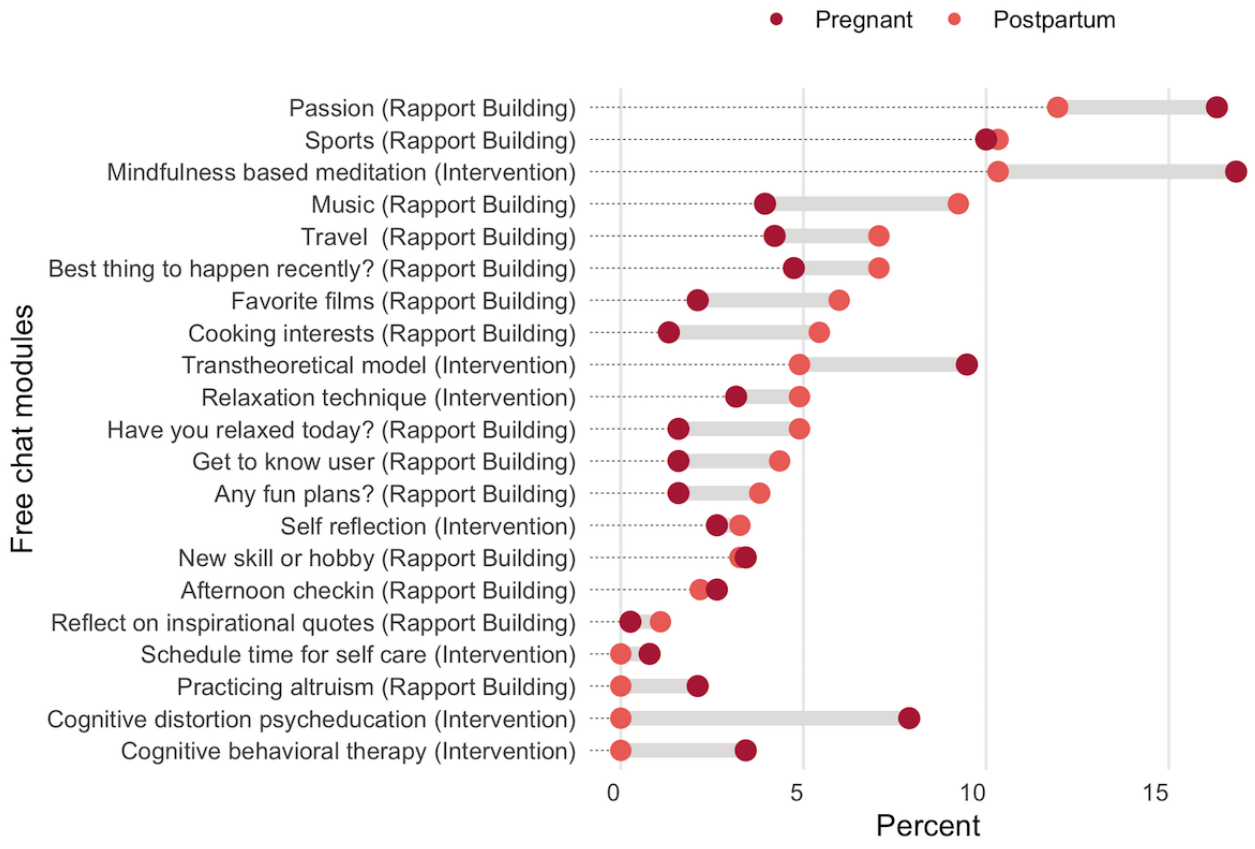
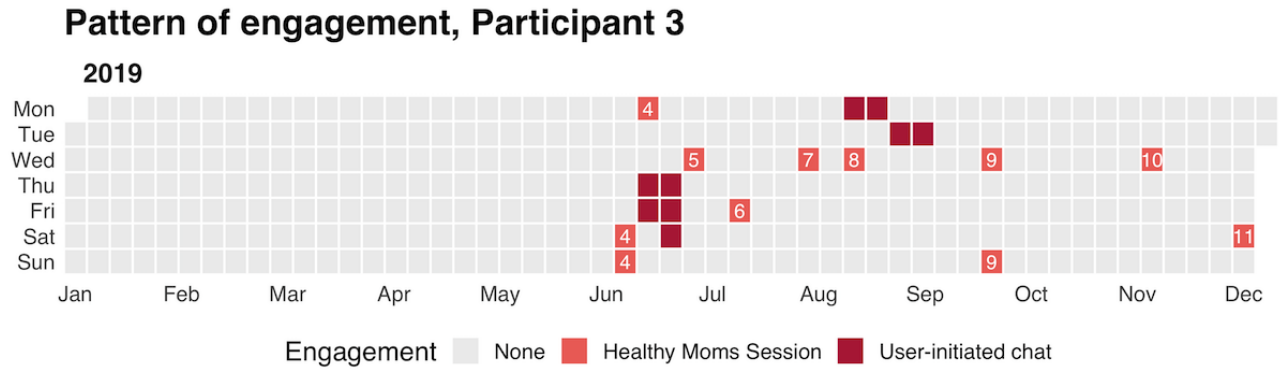


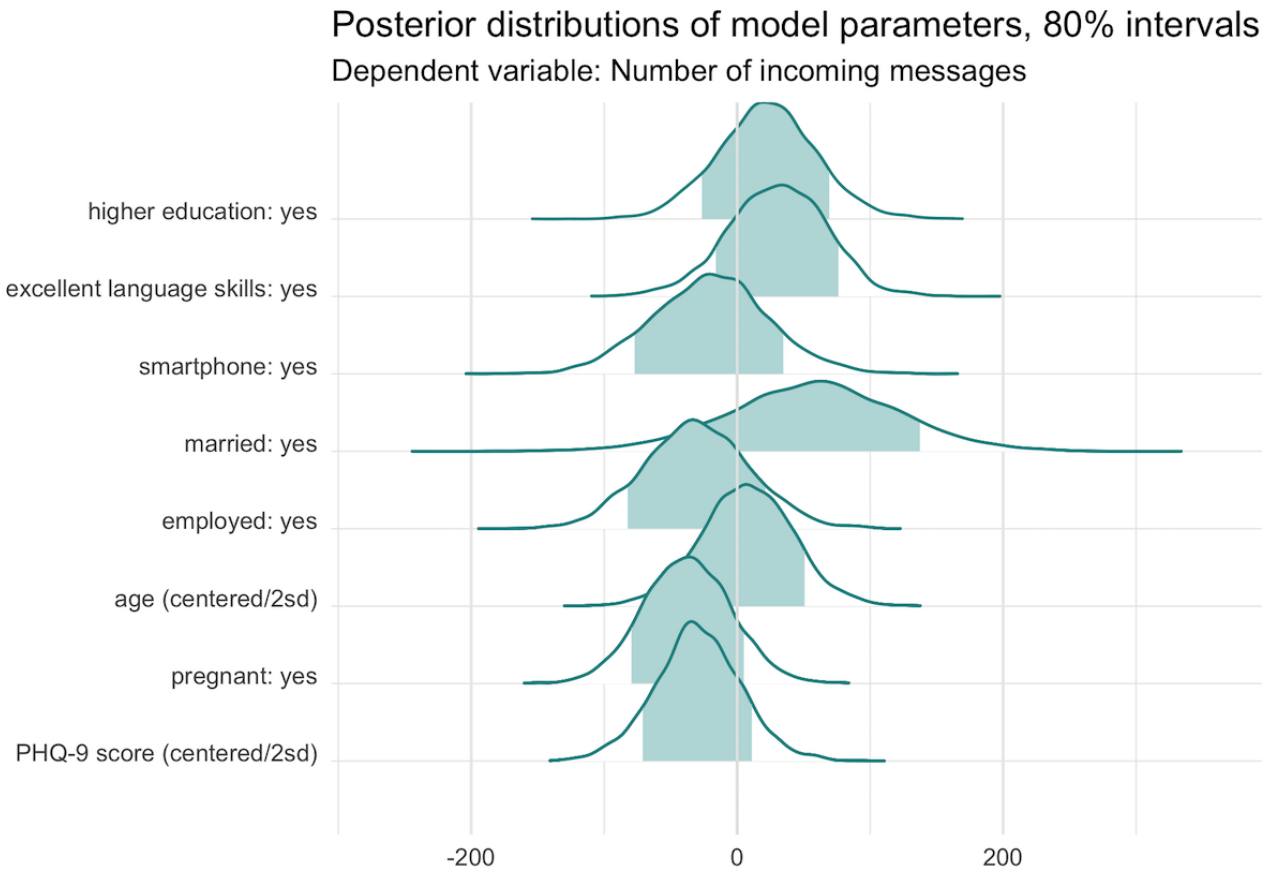
Figure 5. Engagement pattern for Participant 3. Dates shifted to maintain anonymity but pattern preserved.



Correlates of Engagement

To examine the relationship between participant characteristics measured at baseline and intervention engagement, we estimated a Bayesian linear regression model of incoming messages. Figure 6 displays the Markov Chain Monte Carlo draws from the posterior distribution of the parameters. Some evidence suggests that being pregnant (vs a new mom), reporting greater depression symptom severity, and being employed outside of the home are associated with less engagement, whereas being married and more educated are associated with more engagement. For instance, the point estimate is that married women sent 57.8 more messages, holding all else constant. For every 2 SD increase in the baseline PHQ-9 score, holding all else constant, the point estimate is that women sent 29.5 fewer messages.

Figure 6. Results of a Bayesian linear regression model of incoming messages on participants' characteristics measured at baseline (N=40; 1 participant missing required data). The plot shows the Markov Chain Monte Carlo draws from the posterior distribution of the parameters. PHQ-9: Patient Health Questionnaire-9.



Qualitative Findings on Feasibility and Acceptability

Most of the women who were interviewed and who had tried Zuri had a very positive attitude toward the service and expressed that they could trust Zuri. One woman said:

> *It's like a mom to me. My mom is very far, and my sister doesn't have any knowledge of kids.*

Another woman said:

> *I usually keep it to myself. So, when I am chatting with Zuri, it's like they have the right questions to ask*

> *me, and they teach me how to relate with my child, relate with other people.*

Some of the women had also shared Zuri with others, such as their partners or neighbors, who often responded positively. One woman said:

> *My husband was very supportive, because sometimes he used to help me with some answers.*

Many women said that they preferred to chat with Zuri than to chat with a counselor because they felt they could be more open with the automated service. For instance, one woman said:

> *I prefer Zuri because they don't know me.*

Nonetheless, women noticed that Zuri was not perfect and described examples of when Zuri gave an irrelevant response when they asked her a question. Most said they would just ignore the messages and moved on. In our review of chat transcripts, we learned that Zuri was easily confused by messages coming out of order over SMS. This was not an issue on Facebook Messenger; however, almost every woman said they preferred to chat with Zuri through SMS. The main reason being that SMS was free, whereas chatting through Facebook Messenger required them to buy data bundles to access the internet.

Many women mentioned that their favorite part of *Healthy Moms* was the exercises taught by Zuri and the journal, including meditation, breathing, and walking. They found that those exercises were easy and could help them relax. One woman said:

> *They made me be flexible...until my delivery day.*

Other women said that they appreciated the unbiased information provided by Zuri. They indicated that counselors and nurses often give psychosocial advice based on their personal experiences, which can be biased. They felt like they could trust Zuri because she was more unbiased and factual. They especially liked information regarding breastfeeding and how to play with the child. As one woman indicated:

> *For the baby, I never knew she's supposed to be massaged after the bath at all. I never knew she can see different colors.*

Women gave three main reasons why they registered with Zuri and continued to engage. The first reason was anxiety and stress during pregnancy. They were either ashamed of their bodies or worried about experiencing miscarriage. One woman said:

> *One of the negative thoughts I had was maybe if I don't want food what will happen. And then if I sleep bad what will happen to my baby...Actually I was getting worried if I don't feel the movement of my baby inside me sometimes.*

The second reason was that many postpartum women did not feel confident in their roles as new mothers. One woman expressed her anxiety by saying:

> *It's like I don't know how to take care of her, good care of her.*

The final reason was that many of the women interviewed did not have a stable source of income, causing them stress.

Women described 4 main barriers to engaging with Zuri. The first was connectivity. Some women either damaged or lost their phones and did not know how to reconnect with Zuri. The second challenge was that women were easily (and understandably) distracted by their new baby and forgot to complete open sessions. As one woman said:

> *The text can come in the morning, no matter if I am busy or if I am free to answer. If I am free, I just sit and relax. But you see, sometimes we are texting, and the baby starts crying.*

The third challenge was that the registration process was very confusing for some women, especially early on in the study; therefore, some women stopped participating. Related to this, some women were confused by our study's use of 2 SMS short codes: 1 for Zuri and 1 for study assessments. Despite these challenges, women did not contact our study coordinator to receive assistance in using Zuri.

Preliminary Evidence on Response to Treatment

Quantitative Findings

In preparation for modeling the response to treatment, we limited the data to the 12 women who contributed at least 4 mood ratings before and after starting the intervention. Figure 7 plots the time series of ratings by period and overlays the days of intervention engagement with vertical lines.

Figure 8 shows the estimates from a Bayesian linear mixed-effects model. The model included a random effect for observations nested within participants and the following fixed effects: (1) an intercept, (2) a dummy indicator for the treatment phase, (3) a time-within-baseline variable, and (4) a time-within-treatment variable. The time-within-period variables were centered around the first 3 or last 3 observations of the period (first for baseline, last for treatment).

The intercept represents the mean value of the outcome at the first 3 baseline assessments. The treatment indicator is a contrast between the first 3 baseline assessments and last 3 observations in the treatment period, and the time-within-period variables estimate linear change during the baseline and treatment periods.

In this model, the average mood rating at the start of the baseline period was 6.07 on a scale of 0 to 9, and there was no significant baseline trend (an assumption for inference using the multiple baseline design). The point estimate of the treatment effect was 0.42, which represents a 7.0% improvement in mood over the baseline mean (d=0.17). The posterior probability that this effect is greater than zero is 93.2%.

We could not run the same analysis using PHQ-9 scores because we only attempted to collect data at 2 time points and only obtained complete data for half of the (small) sample.

Figure 7. Time series of 432 mood ratings by participant (N=12) and period. Days engaged with Zuri indicated by vertical lines.

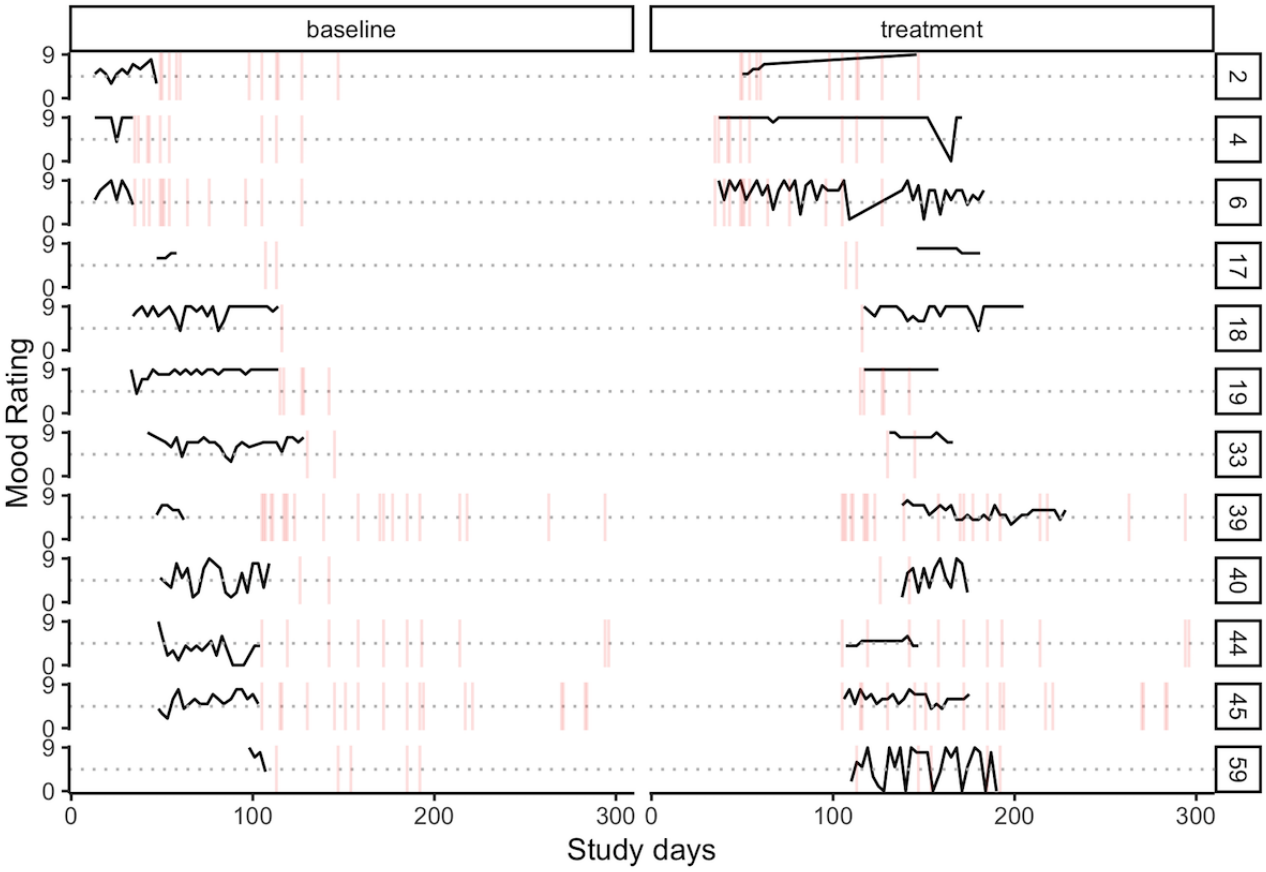
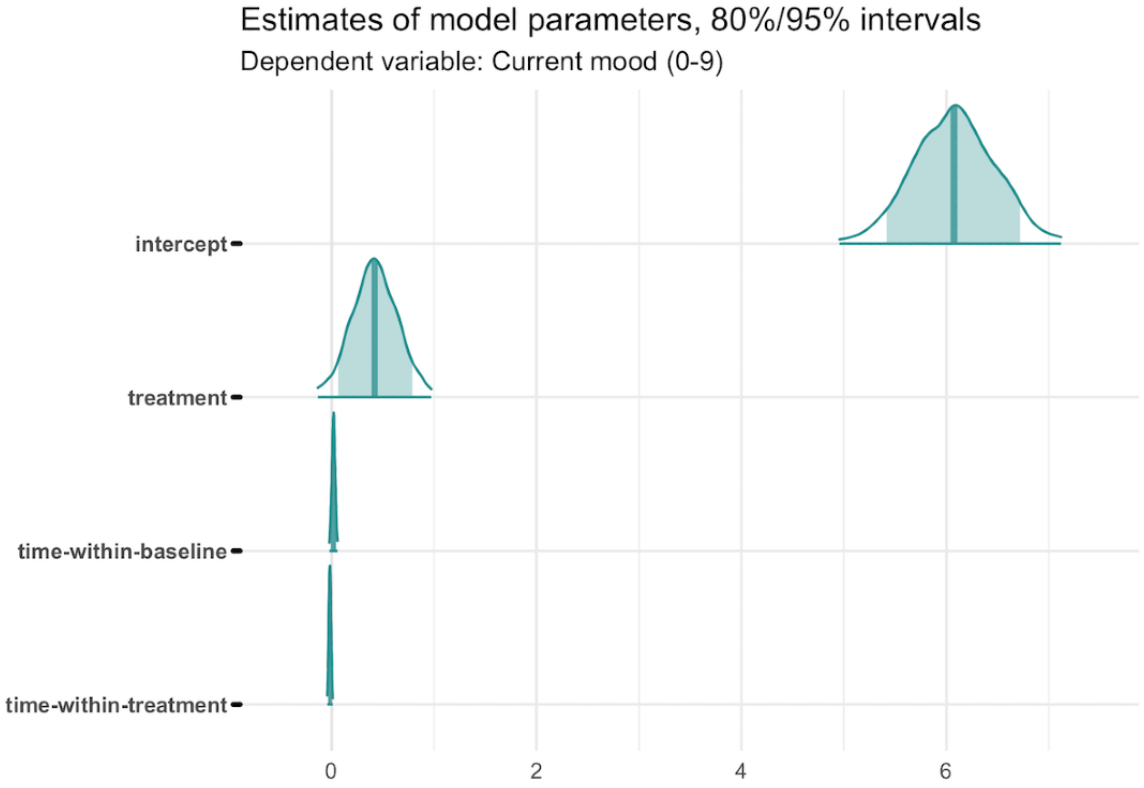


Figure 8. Estimates from a Bayesian linear mixed-effects model of repeated measures data on self-reported mood throughout the study period (432 observations among 12 participants). Uncertainty intervals computed from posterior Markov chain Monte Carlo draws.



Qualitative Findings on Perceived Impact

Many women attributed positive impact to the intervention, which we grouped into 3 themes. The first theme was that Zuri helped them to take care of themselves. Women said that they loved themselves more, that their mood had improved, and that they had learned how to replace negative thoughts with positive thoughts. One woman described her experience with Zuri by saying:

> *Because a pregnant woman is…tired all the time, right? But with Zuri everything was good. I was very active because it also made me have lessons. Because I knew after waking up in the morning I will breathe in and out some minutes. After that I brush, take my breakfast, I wait for noon time something like 12:00 or even 1:00. I go for a walk. After walking I come back shower then I keep myself busy with Zuri. So it's very helpful actually.*

One woman who was ashamed of her body during pregnancy said:

> *I started kind of thinking better, that when you are pregnant, the shape changes and after delivery and doing exercises, everything goes back to normal.*

The second theme was that women acquired new skills that helped them take care of their babies. Many women indicated that they could relate to their child better and experienced less distress raising the child. As one woman said:

> *All those exercises, how to relate to the child, what you do to the child...Honestly, if I hadn't talked to Zuri, I wouldn't know.*

One woman who feared miscarriage even attributed her baby's health and her uncomplicated delivery to Zuri, which we interpret as the woman having found comfort in Zuri during a stressful period.

The last theme was that women experienced improved relationships with others. Some women reported socializing more with others, and this expanding social support system further improved their mood. As one woman said:

> *I used to have the habit of staying alone, not socializing with other people. Zuri made me be able to socialize with people. When they see me doing the exercises, they like knowing where I learnt them from.*

Some women felt more secure and trusted others more. One woman said that she was anxious about leaving her child with another person, even with her family members. However, after finishing a session with Zuri on seeking social support, she explained that she was willing to try asking for help. She reported:

> *So I have tried. [The baby] was comfortable. She cried for some time, then she got used to it.*

Discussion

Principal Results

In this prepilot study, we recruited pregnant women and new mothers in Kenya to try an experimental psychological support service called Zuri. Zuri is a chatbot that engages users in automated, text-based conversations over SMS and Facebook Messenger. Users could initiate chats with Zuri or complete sessions from the *Healthy Moms* perinatal depression intervention curriculum, a CBT-based intervention we adapted from the *Thinking Healthy Program* [16]. We used a single-case experimental design with repeated measures data collection and in-depth interviews to explore the feasibility and acceptability of the service, generate a preliminary estimate of response to treatment, and test study procedures.

Through individual interviews and a review of system logs, we determined that the service was both feasible to deliver and acceptable to this sample of users but not without significant room for improvement and further refinement. Approximately two-thirds of women in the study tried Zuri at least once, and half of those who tried engaged beyond the registration process. This retention rate of 51.9% is slightly more than the average 30-day retention rate of 43% across industries [35] and 40% across provider-prescribed mental health apps specifically [36]. Although our retention rate is based on a small denominator of 27 women who tried the intervention, it suggests that engagement with the initial version of the service is within the range of other digital health apps. Clearly, preventing churn (dropout) is a common challenge.

This was not a clinically referred sample; however, we observed an association between depression severity and intervention use—for every 2 SD increase in the baseline PHQ-9 score, women sent 29.5 fewer messages. This is a small effect in absolute terms; however, it speaks to the potential need for more personalized interventions to maximize user engagement. Most studies on digital mental health apps for common mental disorders such as depression do not report detailed use and usability metrics [37]; however, there is some evidence that also suggests a negative relationship between depression severity and engagement [38].

Users pointed to several positive features of Zuri, including feeling connected to someone who cares while having the benefit of perceived anonymity and privacy of chatting with a machine. This is consistent with existing research showing that people may be more willing to disclose personal information when they believe their responses are not being observed by another person [39], and it probably helps to explain our recruitment experience. More than a quarter of the women who completed the automated screening endorsed having recent suicidal ideation, nearly all of whom accepted our referral to in-person services. Despite having recent and regular contact with antenatal or postpartum medical providers, these women were reporting something to Zuri that they presumably had not reported to frontline medical workers—either because they were not asked, chose not to disclose, or both. There is a substantial latent need for mental health treatment that exists alongside the manifest gaps in access that chatbots such as Zuri could discover and begin to address.

In addition to reporting largely positive impressions of Zuri, users reported modest improvements in mood. To estimate this improvement, we used a multiple baseline design with repeated measures data collection and fit a multilevel model. Importantly,

for making a causal inference, we did not observe an increasing trend in mood during the baseline period. We did, however, observe a small effect in the treatment period. With 432 mood ratings from 12 women before and after beginning the treatment, we estimated that mood improved by 7.0% over the average mood reported at the start of the baseline period (d=0.17). We have high confidence that this effect is greater than zero; however, we are similarly confident that the effect is small. We cannot conclude with confidence that this effect is indeed causal; however, quantifying this effect estimate gives us a benchmark for assessing progress in future iterations of the service. We will look to replicate and hopefully increase this effect in an RCT with a clinically indicated group of users.

We can also look to the digital health and psychotherapy literature for external benchmarks. Although there has been a proliferation of conversational agents for health in recent years [40], the evidence base is small [41]. Two recent RCTs of CBT-based chatbots stand out. In a study of 75 US college students, Tess, an automated chatbot that provides brief psychological interventions over common communication channels such as SMS and Facebook Messenger, reduced the depression symptom severity by roughly 20%, with a reported standardized effect of 0.68 [42]. Another chatbot called Woebot, a stand-alone app that delivers CBT, was tested in a trial with 70 students in the United States. Woebot reduced symptoms of depression by 19%, with a reported standardized effect of 0.44 [43]. For reference, a recent meta-analysis reported that standardized effects of traditional in-person psychotherapy for depression range from 0.66 to 0.77 [44]. Automated conversational agents such as Zuri, Tess, and Woebot have the potential to lower the cost of service delivery while expanding our reach, which could make them highly cost-effective.

Before testing this hypothesis with Zuri, however, we need to build a more robust intervention. As expected with an alpha version, we observed many opportunities for improvement. Some challenges users reported, such as the use of 2 short codes and a confusing registration process, were unique to the setup of this particular study and will not be used again. The bigger challenge will be to make the content more engaging to reduce churn and make the service more robust to misunderstandings. One way to avoid some of the confusion we observed in conversations is to move away from SMS, which can jumble the message order, and instead add a new channel through WhatsApp, the most popular messaging app in Africa [45]. In the short term, this might limit access owing to the cost of internet connectivity; however, penetration rates continue to climb rapidly. From September 2018 to September 2019, the number of data subscriptions in Kenya increased by 23% from 42.2 million to 52 million [46].

In terms of study procedures, we observed a response rate of 13.3% (86/647) among a group of women already enrolled in their county's health SMS program. Seventeen percent of women who completed the screening scored at or above the cutoff for possible depression, and 79% (41/52) of eligible women completed the enrollment process. Depression was not a requirement for inclusion in this study; however, it will be in future studies. Our experience in this prepilot study suggests an overall enrollment rate of 1%, taking depression symptoms into account. Therefore, to recruit a sample of 100 possibly depressed pregnant women and new mothers in a future trial using the same remote procedures, these estimates suggest that we would need to advertise to a pool of at least 10,000 women. This can be easily achieved through print and digital advertising. In Nairobi County alone, there were more than 130,000 live births in 2017 [47].

Our experience with remote automated data collection suggests that women were willing and able to reply to a 1-question prompt asking them to rate their current mood. However, we were less successful at obtaining end-line data using the PHQ-9. In a future trial, it will be important to budget and plan for study staff to augment remote data collection procedures.

Limitations

The objective of this prepilot study was to adapt *Thinking Healthy* for delivery through Zuri for developing and testing study procedures to inform the design of a future trial and to generate preliminary evidence to guide the next round of Zuri's development. We were limited in our pursuit of these objectives given that we only offered screening and conversations in English. This likely constrained our recruitment efforts as non–English-speaking women did not have the opportunity to participate. This implies that our estimates for future recruitment are conservative. The other main limitation of operating Zuri in English is that we do not have data on how Zuri functions in Swahili. This is a priority target for development. A related limitation is that, by virtue of requiring advanced language skills, we recruited a highly educated sample of women relative to the general population. In a future trial, it will be important to explore how women of all educational backgrounds engage with Zuri.

Conclusions

We determined that Zuri is feasible to deliver via SMS and acceptable to a sample of pregnant women and new mothers recruited from 2 large public hospitals in Kenya. The results of this prepilot study will serve as a baseline for future studies in terms of recruitment, data collection, and outcomes. The next step in Zuri's development is to refine the intervention content and add Swahili language support. Conversational agents such as Zuri have great potential to address the large treatment gap that exists in many low-resource settings, both as a new channel of treatment and as an adjunct to traditional and task-shifting approaches.

Acknowledgments

EG, NP, SR, MR, and AJ designed the study. All authors provided input on the design and planning and contributed to the study implementation. Funding was provided by the Duke Global Health Institute and Bass Connections at Duke University. The

sponsors had no input in the process of data collection, analysis, or interpretation. The authors thank Anna-Karin Hess and Lulla Kiwinda for formative work on Zuri and Josh Brown for his technical assistance with Tess.

Conflicts of Interest

MR is the CEO and Founder of X2AI and created Tess. AJ is an employee of X2AI. EG is an unpaid advisor to the X2AI Ethical Advisory Board and has no financial stake in the company.

Multimedia Appendix 1
Example Healthy Moms session.
[PDF File (Adobe PDF File), 63 KB - formative_v4i10e17895_app1.pdf]

Multimedia Appendix 2
Journal content related to example Healthy Moms session.
[PDF File (Adobe PDF File), 1920 KB - formative_v4i10e17895_app2.pdf]

References

1. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005 Nov;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
2. Villegas L, McKay K, Dennis C, Ross L. Postpartum depression among rural women from developed and developing countries: a systematic review. J Rural Health 2011;27(3):278-288. [doi: 10.1111/j.1748-0361.2010.00339.x] [Medline: 21729155]
3. Khalifeh H, Hunt IM, Appleby L, Howard LM. Suicide in perinatal and non-perinatal women in contact with psychiatric services: 15 year findings from a UK national inquiry. Lancet Psychiatry 2016 Mar;3(3):233-242. [doi: 10.1016/S2215-0366(16)00003-1] [Medline: 26781366]
4. Oates M. Perinatal psychiatric disorders: a leading cause of maternal morbidity and mortality. Br Med Bull 2003;67:219-229. [doi: 10.1093/bmb/ldg011] [Medline: 14711766]
5. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: a review. Infant Behav Dev 2010 Feb;33(1):1-6 [FREE Full text] [doi: 10.1016/j.infbeh.2009.10.005] [Medline: 19962196]
6. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. Lancet Psychiatry 2016 Oct;3(10):973-982 [FREE Full text] [doi: 10.1016/S2215-0366(16)30284-X] [Medline: 27650773]
7. Grigoriadis S, VonderPorten EH, Mamisashvili L, Tomlinson G, Dennis C, Koren G, et al. The impact of maternal depression during pregnancy on perinatal outcomes: a systematic review and meta-analysis. J Clin Psychiatry 2013 Apr;74(4):e321-e341. [doi: 10.4088/JCP.12r07968] [Medline: 23656857]
8. Rahman A, Hafeez A, Bilal R, Sikander S, Malik A, Minhas F, et al. The impact of perinatal depression on exclusive breastfeeding: a cohort study. Matern Child Nutr 2016 Jul;12(3):452-462 [FREE Full text] [doi: 10.1111/mcn.12170] [Medline: 25682731]
9. Surkan PJ, Patel SA, Rahman A. Preventing infant and child morbidity and mortality due to maternal depression. Best Pract Res Clin Obstet Gynaecol 2016 Oct;36:156-168. [doi: 10.1016/j.bpobgyn.2016.05.007] [Medline: 27422745]
10. Beck CT. The effects of postpartum depression on child development: a meta-analysis. Arch Psychiatr Nurs 1998 Feb;12(1):12-20. [doi: 10.1016/s0883-9417(98)80004-6] [Medline: 9489170]
11. Gentile S. Untreated depression during pregnancy: short- and long-term effects in offspring. A systematic review. Neuroscience 2017 Feb 7;342:154-166. [doi: 10.1016/j.neuroscience.2015.09.001] [Medline: 26343292]
12. Junge C, Garthus-Niegel S, Slinning K, Polte C, Simonsen TB, Eberhard-Gran M. The impact of perinatal depression on children's social-emotional development: a longitudinal study. Matern Child Health J 2017 Mar;21(3):607-615. [doi: 10.1007/s10995-016-2146-2] [Medline: 27485491]
13. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to prevent perinatal depression: evidence report and systematic review for the US preventive services task force. J Am Med Assoc 2019 Feb 12;321(6):588-601. [doi: 10.1001/jama.2018.20865] [Medline: 30747970]
14. Lund C, Tomlinson M, de Silva M, Fekadu A, Shidhaye R, Jordans M, et al. PRIME: a programme to reduce the treatment gap for mental disorders in five low- and middle-income countries. PLoS Med 2012;9(12):e1001359 [FREE Full text] [doi: 10.1371/journal.pmed.1001359] [Medline: 23300387]
15. Baron EC, Hanlon C, Mall S, Honikman S, Breuer E, Kathree T, et al. Maternal mental health in primary care in five low- and middle-income countries: a situational analysis. BMC Health Serv Res 2016 Feb 16;16:53 [FREE Full text] [doi: 10.1186/s12913-016-1291-z] [Medline: 26880075]

16. <p/>. Thinking Healthy: A Manual for Psychological Management of Perinatal Depression. World Health Organization. 2015. URL: https://www.who.int/mental_health/maternal-child/thinking_healthy/en/ [accessed 2019-03-29] [WebCite Cache ID 77F8iMHud]
17. Rahman A, Malik A, Sikander S, Roberts C, Creed F. Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: a cluster-randomised controlled trial. Lancet 2008 Sep;372(9642):902-909. [doi: 10.1016/s0140-6736(08)61400-2]
18. Baranov V, Bhalotra S, Biroli P, Maselko J. Maternal depression, women's empowerment, and parental investment: evidence from a randomized controlled trial. Am Econ Rev 2020 Mar 1;110(3):824-859 [FREE Full text] [doi: 10.1257/aer.20180511]
19. Fuhr DC, Weobong B, Lazarus A, Vanobberghen F, Weiss HA, Singla DR, et al. Delivering the thinking healthy programme for perinatal depression through peers: an individually randomised controlled trial in India. Lancet Psychiatry 2019 Feb;6(2):115-127. [doi: 10.1016/S2215-0366(18)30466-8] [Medline: 30686385]
20. Rahman A. Challenges and opportunities in developing a psychological intervention for perinatal depression in rural Pakistan--a multi-method study. Arch Womens Ment Health 2007;10(5):211-219 [FREE Full text] [doi: 10.1007/s00737-007-0193-9] [Medline: 17676431]
21. Maselko J, Sikander S, Bhalotra S, Bangash O, Ganga N, Mukherjee S, et al. Effect of an early perinatal depression intervention on long-term child development outcomes: follow-up of the thinking healthy programme randomised controlled trial. Lancet Psychiatry 2015 Jul;2(7):609-617. [doi: 10.1016/S2215-0366(15)00109-1] [Medline: 26303558]
22. Padmanathan P, de Silva MJ. The acceptability and feasibility of task-sharing for mental healthcare in low and middle income countries: a systematic review. Soc Sci Med 2013 Nov;97:82-86. [doi: 10.1016/j.socscimed.2013.08.004] [Medline: 24161092]
23. Watson PJ, Workman EA. The non-concurrent multiple baseline across-individuals design: an extension of the traditional multiple baseline design. J Behav Ther Exp Psychiatry 1981 Sep;12(3):257-259. [doi: 10.1016/0005-7916(81)90055-0] [Medline: 7320215]
24. Green EP, Pearson N, Rajasekharan S, Rauws M, Joerin A, Kwobah E, et al. Expanding access to depression treatment in Kenya through automated psychological support: protocol for a single-case experimental design pilot study. JMIR Res Protoc 2019 Apr 29;8(4):e11800 [FREE Full text] [doi: 10.2196/11800] [Medline: 31033448]
25. Sheppard E. How WhatsApp and SMS Are Being Used to Save the Lives of Babies in Africa. The Guardian. 2018. URL: https://www.theguardian.com/business-call-to-action-partnerzone/2018/aug/09/how-whatsapp-and-sms-are-being-used-to-save-the-lives-of-babies-in-africa [accessed 2019-02-23] [WebCite Cache ID 76PXGUjT3]
26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
27. Green E. The Healthy Moms Team. Healthy Moms 2020:- [FREE Full text] [doi: 10.17605/OSF.IO/4KPZ2]
28. Green E. Get to Know Our Pop-Up UX Lab in Nairobi. Medium. 2018. URL: https://medium.com/@ericpgreen/healthymoms-a4c573fbfb7f [accessed 2018-08-02] [WebCite Cache ID 71NU8adLb]
29. Glaser BG. The constant comparative method of qualitative analysis. Soc Prob 1965 Apr;12(4):436-445. [doi: 10.1525/sp.1965.12.4.03a00070]
30. Rindskopf D. Bayesian analysis of data from single case designs. Neuropsychol Rehabil 2014;24(3-4):572-589. [doi: 10.1080/09602011.2013.866903] [Medline: 24365037]
31. Bürkner P. An R package for bayesian multilevel models using Stan. J Stat Soft 2017;80(1):- [FREE Full text] [doi: 10.18637/jss.v080.i01]
32. <p/>. Security. Mental Health Chatbot. 2019. URL: https://www.x2ai.com/security [accessed 2019-02-23] [WebCite Cache ID 76PYM33We]
33. Green E. Zuri Pre-pilot Repository. GitHub. 2020. URL: https://github.com/ericpgreen/zuri-2019-stage2rr [accessed 2020-09-10]
34. Green EP, Tuli H, Kwobah E, Menya D, Chesire I, Schmidt C. Developing and validating a perinatal depression screening tool in Kenya blending Western criteria with local idioms: a mixed methods study. J Affect Disord 2018 Mar 1;228:49-59. [doi: 10.1016/j.jad.2017.11.027] [Medline: 29227955]
35. Perro J. Mobile Apps: What's A Good Retention Rate? Upland Localytics. URL: http://info.localytics.com/blog/mobile-apps-whats-a-good-retention-rate; [accessed 2020-09-10]
36. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. IQVIA. 2015. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2020-09-22]
37. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
38. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]

39. Lucas G, Gratch J, King A, Morency L. It's only a computer: virtual humans increase willingness to disclose. Comput Hum Behav 2014 Aug;37:94-100 [FREE Full text] [doi: 10.1016/j.chb.2014.04.043]
40. Montenegro JL, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
41. Laranjo L, Dunn A, Tong H, Kocaballi A, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
42. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
43. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (WOEBOT): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
44. Cuijpers P, Karyotaki E, Reijnders M, Huibers MJ. Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. Cogn Behav Ther 2018 Mar;47(2):91-106. [doi: 10.1080/16506073.2017.1420098] [Medline: 29345530]
45. Dahir A. WhatsApp is the Most Popular Messaging App in Africa. Quartz. 2018. URL: https://qz.com/africa/1206935/whatsapp-is-the-most-popular-messaging-app-in-africa/; [accessed 2020-09-10]
46. <p/>. First Quarter Sector Statistics Report for the Financial Year 2019/2020 (July - September 2019). Communications Authority of Kenya. 2020. URL: https://ca.go.ke/wp-content/uploads/2019/12/Sector-Statistics-Report-Q1-2019-2020.pdf [accessed 2020-09-10]
47. Murphy GA, Waters D, Ouma PO, Gathara D, Shepperd S, Snow RW, et al. Estimating the need for inpatient neonatal services: an iterative approach employing evidence and expert consensus to guide local policy in Kenya. BMJ Glob Health 2017;2(4):e000472 [FREE Full text] [doi: 10.1136/bmjgh-2017-000472] [Medline: 29177099]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
LMIC: low- and middle-income country
mhGAP: Mental Health Gap Action Programme
PHQ-9: Patient Health Questionnaire–9
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.01.20; peer-reviewed by J Duffecy, D Singla, C Stiles-Shields; comments to author 12.03.20; revised version received 09.08.20; accepted 01.09.20; published 05.10.20.

Please cite as:
Green EP, Lai Y, Pearson N, Rajasekharan S, Rauws M, Joerin A, Kwobah E, Musyimi C, Jones RM, Bhat C, Mulinge A, Puffer ES Expanding Access to Perinatal Depression Treatment in Kenya Through Automated Psychological Support: Development and Usability Study
JMIR Form Res 2020;4(10):e17895
URL: https://formative.jmir.org/2020/10/e17895
doi:10.2196/17895
PMID:33016883

Original Paper

Ethnicity Differences in Sleep Changes Among Prehypertensive Adults Using a Smartphone Meditation App: Dose-Response Trial

John C Sieverdes[1], PhD; Frank A Treiber[2,3], PhD; Christopher E Kline[4], PhD; Martina Mueller[2], PhD; Brenda Brunner-Jackson[3], MPH; Luke Sox[2], BSc; Mercedes Cain[2], BSc; Maria Swem[2], BSc; Vanessa Diaz[3], MD; Jessica Chandler[2], PhD

[1]College of Charleston, Health and Human Performance, Charleston, SC, United States

[2]College of Nursing, Medical University of South Carolina, Charleston, SC, United States

[3]College of Medicine, Medical University of South Carolina, Charleston, SC, United States

[4]Department of Health & Physical Activity, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
John C Sieverdes, PhD
College of Charleston
Health and Human Performance
24 George Street
Charleston, SC
United States
Phone: 1 843 953 6094
Email: sieverdesjc@cofc.edu

Abstract

Background: African Americans (AAs) experience greater sleep quality problems than non-Hispanic Whites (NHWs). Meditation may aid in addressing this disparity, although the dosage levels needed to achieve such benefits have not been adequately studied. Smartphone apps present a novel modality for delivering, monitoring, and measuring adherence to meditation protocols.

Objective: This 6-month dose-response feasibility trial investigated the effects of a breathing awareness meditation (BAM) app, Tension Tamer, on the secondary outcomes of self-reported and actigraphy measures of sleep quality and the modulating effects of ethnicity of AAs and NHWs.

Methods: A total of 64 prehypertensive adults (systolic blood pressure <139 mm Hg; 31 AAs and 33 NHWs) were randomized into 3 different Tension Tamer dosage conditions (5,10, or 15 min twice daily). Sleep quality was assessed at baseline and at 1, 3, and 6 months using the Pittsburgh Sleep Quality Index (PSQI) and 1-week bouts of continuous wrist actigraphy monitoring. The study was conducted between August 2014 and October 2016 (IRB #Pro00020894).

Results: At baseline, PSQI and actigraphy data indicated that AAs had shorter sleep duration, greater sleep disturbance, poorer efficiency, and worse quality of sleep (range P=.03 to P<.001). Longitudinal generalized linear mixed modeling revealed a dose effect modulated by ethnicity (P=.01). Multimethod assessment showed a consistent pattern of NHWs exhibiting the most favorable responses to the 5-min dose; they reported greater improvements in sleep efficiency and quality as well as the PSQI global value than with the 10-min and 15-min doses (range P=.04 to P<.001). Actigraphy findings revealed a consistent, but not statistically significant, pattern in the 5-min group, showing lower fragmentation, longer sleep duration, and higher efficiency than the other 2 dosage conditions. Among AAs, actigraphy indicated lower sleep fragmentation with the 5-min dose compared with the 10-min and 15-min doses (P=.03 and P<.001, respectively). The 10-min dose showed longer sleep duration than the 5-min and 15-min doses (P=.02 and P<.001, respectively). The 5-min dose also exhibited significantly longer average sleep than the 15-min dose (P=.03).

Conclusions: These findings indicate the need for further study of the potential modulating influence of ethnicity on the impact of BAM on sleep indices and user-centered exploration to ascertain the potential merits of refining the Tension Tamer app with attention to cultural tailoring among AAs and NHWs with pre-existing sleep complaints.

(JMIR Form Res 2020;4(10):e20501) doi:10.2196/20501

KEYWORDS

meditation; sleep; mobile phone; prehypertension; ethnicity

Introduction

Background

Obtaining an adequate amount and quality of sleep is essential for optimal health. Chronically insufficient sleep negatively impacts daytime function and has been linked to an increased risk of hypertension, stroke, cardiovascular disease, obesity, stress, and poor mood [1-6]. One of the most common contributors to poor sleep is chronic stress [7-11], which can disrupt sleep by increasing rumination, worries, intrusive thoughts, and heightened muscular tension [12,13]. Mind-body strategies addressing stress may not only aid in the improvement of sleep quality but also reduce exacerbation of other chronic conditions such as hypertension, cardiovascular disease, diabetes, and obesity [14-16].

Mitigating stress through positive cognitive behavioral coping strategies is a common treatment approach to improve sleep quality [11]. Meditation is one tactic that can be incorporated into a comprehensive treatment plan among those with chronic stress experiencing trouble sleeping [8,10]. A variety of meditation approaches already exist, including transcendental meditation (TM) and breathing awareness meditation (BAM). There is evidence to show improvements in anxiety, stress [17-19], and blood pressure (BP) through the use of these types of meditation techniques [20-23]. These practices focus on relaxation, self-acceptance, and staying in the present moment through slow, diaphragmatic breathing.

Although the benefits of meditation on stress are positive, sleep outcomes have been mixed. Neuendorf et al [11] reviewed 112 mind-body–based randomized controlled trials (RCTs) on sleep quality outcomes. Of the 23 meditation RCTs, only 39% reported beneficial effects on sleep quality. A 2016 meta-analysis of 6 RCTs using mindfulness meditation incorporating BAM reported statistically significant but mild improvements in self-reported total wake time, sleep onset latency, sleep efficiency, and sleep quality [24]. Another meta-analysis of 47 RCTs found minimal evidence of meditation studies improving positive mood, attention, eating habits, or sleep [17]. Collectively, the findings are somewhat mixed, with mild-to-moderate effects of meditation interventions on sleep quality. Potential reasons for the variability in outcomes included heterogeneity of participants, with some studies using individuals identified as having sleep disorders, some studies using patients with mental health disorders associated with sleep disturbances (eg, anxiety, depression, posttraumatic stress disorder [PTSD]), and others assessing sleep outcomes among individuals not recruited for sleep issues. Many of these studies were conducted for a short term (eg, 8-12 weeks), often relied upon self-report rather than multimethod assessment (eg, inclusion of polysomnography and actigraphy), did not objectively monitor adherence to meditation protocols, and failed to report study details such as the level of expertise of the trainers and the degree of blinding of the evaluators. [11,13,24]. Furthermore, none of the RCTs reviewed were designed to evaluate dose-response effects on sleep disturbance indices.

Large-scale dissemination of meditation programs using traditional modes of training may be challenging. Individual or group-based in-person training and monitoring methodologies have typically been used [11,17]; however, these approaches often result in challenges of travel, child care, and other logistical barriers that may reduce participation and retention rates. Finally, not having an objective indicator of adherence to meditation sessions outside of the in-person sessions has been noted as a potential contributor to the variability in outcomes across studies [11,13,24].

Technology-delivered meditation programs are an innovative option to overcome several of these barriers. Unfortunately, few programs have been rigorously evaluated. In particular, meditation programs delivered by smartphone apps are widely available [25] and can be used for reducing stress and for other clinical uses such as corresponding reductions in BP and improved sleep. However, the effects of these meditation apps on sleep indices have received sparse empirical evaluation [26].

Research has also consistently shown greater self-reported sleep quality problems (eg, longer latency of onset, shorter duration, greater disruption during sleep) among African Americans (AAs) than non-Hispanic Whites (NHWs) [3,27-31]. These findings have typically been supported in studies that included both objective (eg, wrist actigraphy, polysomnography) and self-reported measures [32-35]. The Tension Tamer study's use of self-reported and actigraphy-based sleep assessment will aid in determining whether ethnic differences in sleep quality are observed via both self-reported and objective measurements. The available data enable a preliminary exploration of whether ethnicity potentially modulates prehypertensive adult responsivity to varying doses of the Tension Tamer app upon sleep.

Further studies are warranted to determine if meditation techniques, such as BAM, have a beneficial dose-response influence upon indices of sleep based on self-reported and objective verification via accelerometry. A recently completed 6-month dose-response study using Tension Tamer, a BAM smartphone app, provides this opportunity. The primary goal of the feasibility trial was to determine the potential dose-response impact on resting BP among prehypertensive adults [23]. We examined the reduction of resting systolic BP (SBP) compared with baseline levels for the 3 twice-daily doses (5, 10, and 15 min) of using the app at each evaluation point (1, 3, and 6 months). All dosage conditions had statistically significant reductions in BP, with 5-min and 10-min conditions showing average reductions in SBP (−5 to −8 mm Hg) comparable with other BAM and TM trials with prehypertensive and hypertensive samples [23,36-41]. The 15-min dose condition showed larger average SBP reductions compared with the 5-min and 10-min dose conditions, ranging from −10.9 mm Hg to −12.4 mm Hg across the 6-month trial. Secondary analyses are now directed at evaluating the potential impact of varying BAM doses on known modulators of BP control, such as sleep quality [42-44].

Objectives

This paper reports on the exploratory analysis of 3 different doses of BAM on sleep quality changes in adult NHWs and

AAs with prehypertension. We hypothesized that increased doses of BAM would impart greater improvements in self-reported and actigraphy-derived measures of sleep quality. We also hypothesized that AAs would exhibit poorer baseline levels of sleep quality compared with NHWs.

Methods

Participants

A total of 287 prehypertensive adults were recruited using paper advertising and word-of-mouth referrals at an academic health center in southeastern United States. Study candidates were contacted by phone to initiate eligibility screening and were scheduled for the first of 3 clinic BP evaluations. Inclusion criteria included participants aged between 18 and 90 years; being NHWs or AAs; and having systolic prehypertension based on the Seventh Report of the National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (120 mm Hg-139 mm Hg) on 3 consecutive visits, each at least 2 days apart. Exclusion criteria included unwillingness to accept randomization into 1 of 3 dosage groups, current medication regimens that affected BP, inability to open and navigate the Tension Tamer smartphone app after a hands-on training, or inability of participants to position their fingers over their smartphone camera because of tremors. Additional exclusion criteria included being currently pregnant, lactating, or having intention to become pregnant during the trial, participating in another research study of any kind, inability to read or speak English, and not having smartphone signal at the participant's place of residence. The study was conducted between August, 2014 and October, 2016.

Description of Tension Tamer App

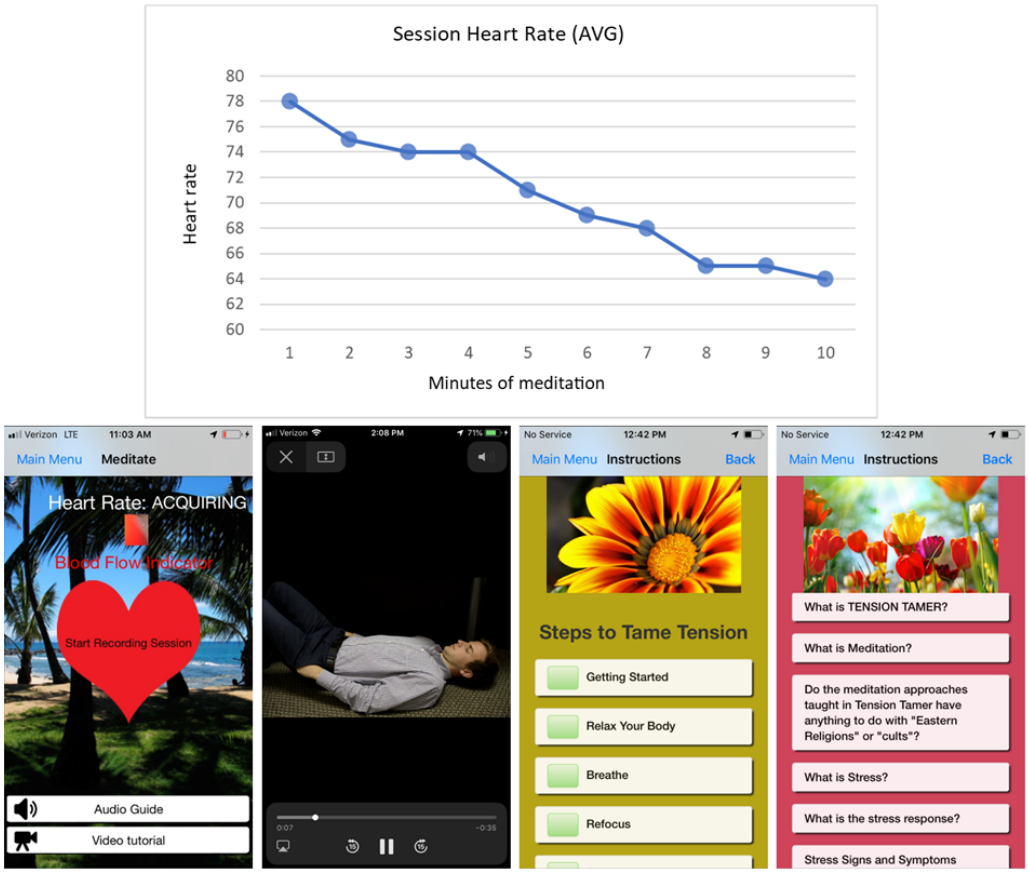
Tension Tamer captures real-time heart rate (HR) with a proprietary software involving reflective photoplethysmography by using the smartphone video camera and light-emitting diode (LED) flashlight [45]. The video camera and software can detect HR pulses through the finger's capillary beds and output the digital reading on the interface and store the session HR data. This feature has been previously validated against electrocardiogram-derived HR (Pearson r≥.99, standard error of the estimate≤2.09 bpm) [45]. Graphical HR feedback is provided after each Tension Tamer session to illustrate the average HR changes across each minute (Figure 1).

Users have the option to turn off (or reactivate) the audio BAM instructions, set session midway and ending alerts (chime or gong), and select different background themes for screenshots (eg, nature, rock garden, beach scenes; Figure 1). The Tension Tamer app sends duration and time stamp–encrypted HR data to our institution's Health Insurance Portability and Accountability Act–compliant server for storage in a relational database management system via a securely authenticated web interface. The time-stamped HR data provide the duration of session data, which is used to facilitate the measurement of session adherence.

The Tension Tamer app was developed and tailored for prehypertensive adults using a patient-centered iterative design approach guided by behavioral change theories. These included the self-determination theory [46] and the social cognitive theory [47] to guide the development of the behavioral content and the *what*, *how*, and *why* of Tamer Tension engagement and maintenance by fostering self-efficacy and intrinsic motivation to sustain engagement in the meditation sessions over time [48]. The People, Activity, Context, and Technology (PACT) model guided the presentation and engagement process of using the app. The PACT model posits that users must feel at ease with and perceive the technology as useful in reaching a desired goal [49]. Feedback based on the levels of adherence in the Tension Tamer program was used to provide social reinforcement and motivation via SMS messages to increase self-efficacy and autonomous motivation to increase adoption and sustain engagement of Tension Tamer for the management of stress and BP. As noted earlier, the primary outcomes of changes in BP from the 6-month dose-response trial involving prehypertensive adults have been reported [23].

Sleep data were collected through self-reported surveys and 1-week continuous actigraphy acquisition at each evaluation time point (baseline and months 1, 3, and 6). All research procedures were approved by the institutional review board before conducting the study (IRB #Pro00020894).

Figure 1. The graphical heart rate feedback after each Tension Tamer session to illustrate average heart rate changes across each minute using the smartphone's camera (top). Screenshots of the home screen (bottom), breathing meditation instructions, information regarding how Tension Tamer works, and a frequently asked questions page.



Procedures

Complete details of the study protocol have been reported elsewhere [23]. After informed consent was obtained, and 3 separate confirmatory BP evaluations were done to verify the prehypertensive status, baseline measures of demographics (ie, ethnicity, education, income, height, weight, and physical activity level) were taken and a set of questionnaires including the Pittsburgh Sleep Quality Index (PSQI) were administered [50]. BMI was calculated as weight in kilograms divided by height in meters squared. Participants also wore an ActiGraph monitor for 7 consecutive 24-hour periods. Participants were subsequently randomized into 1 of 3 Tension Tamer dosage conditions of 5, 10, or 15 min, each to be performed twice daily.

The Tension Tamer app was installed on either the iOS (iPhone operating system) or the Android operating system on each participant's personal smartphone. If the subject did not have a smartphone, an Apple iPhone 4 or 5 or Samsung Galaxy 4 or 5 was provided at no cost for the duration of the study. A staff member trained in the use of the Tension Tamer app was present to provide guidance as needed during the introductory tutorial. The Tension Tamer app tutorial included Tension Tamer videos explaining BAM, demonstrating diaphragmatic breathing, and explaining how to operate the app's features, including obtaining video camera HR capture using the fingertip. An audio recording of a British female voice guided them through their first BAM session. Participants were instructed to place the tip of a finger over the camera's lens and LED light. The app would vibrate if the finger were misaligned and prompt for realignment. Participants were directed in the audio guide to sit comfortably, close their eyes, and attend to the movement of their diaphragm while taking deep, slow, *belly* breaths. At the start of each session, a chime indicated the start with an on-screen timer. Additional chimes notified the user of the midpoint and the end of the session. Numeric feedback was immediately displayed on a 4-beat rolling average basis throughout the session. A graph was displayed immediately following the session, showing the average HR response during each minute of the session (shown in Figure 1). The training session ended with the participant demonstrating the ability to initiate another Tension Tamer session, including HR acquisition, before returning home. In subsequent BAM sessions, the participants could toggle the guided audio training on or off as desired.

An ActiGraph monitor was then fitted on their wrist, and participants were given instructions to wear the device for 7 consecutive days and to complete a daily sleep log. The same procedure was followed at each subsequent laboratory evaluation

at 1, 3, and 6 months. Participants were compensated US $145 for their time after completion of each evaluation.

Measures

Actigraphy

Actigraphy was captured using the ActiSleep monitor (ActiGraph models GT3XP-BTLE and ASP-BTLE; ActiGraph Corp). The ActiSleep monitor is a research-grade accelerometer that provides objective estimates of physical movement and includes multiple sleep parameters that have been validated against polysomnography [51-54]. The wrist monitor was secured on the nondominant hand using the manufacturer's Velcro wrist strap. Participants were told to wear the device at all times except during activities that would get the device wet (eg, showering, bathing, swimming). Epochs were set to 10-second intervals. Participants were given sleep logs for recording each night's sleep time and wake time and to note any disturbances (eg, awakened by noise, having bad dreams, using the bathroom). Data were downloaded and processed using ActiLife 6 software (version 6.13.3, ActiGraph LLC), which converted the data to 60-second epochs. Initially, the software's automated sleep detection algorithm was used to identify the sleep periods. Comparisons were performed against the sleep log data by a doctoral-level analyst, and adjustments were made when the sleep and wake times differed by more than 15 min. Sleep and wake changes were identified using visual examination of the graphical output, self-reported notes on sleep logs, nonwear periods, lux readouts from the light sensor, and patterns across all days. The Cole-Kripke algorithm [55] was used to categorize the sleep and wake status during each sleep period. The following variables were averaged at each time point and retained for analysis: sleep time per night (in min), sleep efficiency (ie, the percentage of time in bed that was spent asleep), and sleep fragmentation index (ie, an ActiSleep-derived index of restlessness during the entire sleep period by measurement of the amount of movement while in bed; lower values indicate better sleep). At least 4 days of valid data, including 1 weekend night, were required at each evaluation to be included in the analyses.

PSQI

PSQI, administered at each evaluation, is a well-validated self-report questionnaire of sleep quality over the previous month [50,56-59]. The 19 items on PSQI use both open-ended (eg, regular bedtime and awakening times) and 5-point Likert scale formats. Scoring algorithms provide 7 components (subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbances, sleep medication, and daytime dysfunction). The total sleep quality score (0-21) is derived from the 7 components (each scored 0-3). Higher scores on each component are indicative of poorer sleep quality. In addition to the 7 components, Cole et al [53] identified 3 latent factors based on a combination of PSQI components: sleep efficiency (sleep duration and sleep efficiency), perceived sleep quality (subjective sleep quality, sleep latency, and sleep medication), and daily disturbances (sleep disturbances and daytime dysfunction). A recent meta-analysis of 37 PSQI psychometric studies has revealed good internal consistency (Cronbach alpha=.70-.83), construct validity (ie, discriminated between groups with a disorder associated with poor sleep, such as depression and PTSD, and healthy individuals), convergent validity (eg, moderate positive correlations with anxiety and depression), and divergent validity (eg, weak associations with unrelated constructs such as vomiting and anger) [58]. The 3-factor structure has been replicated across a variety of populations involving men and women with various chronic illnesses or conditions (eg, depression, pregnant women, kidney transplant recipients, PTSD, perimenopause, early postmenopausal women) as well as those without any known diseases. Participants have represented multiple ethnic groups (eg, AAs, NHWs, Hispanic Americans) [59-62]. We used the 3 factors (sleep efficiency, sleep quality, and daily disturbances) along with the global score in the statistical analyses.

Physical Activity

In addition to the objective physical activity data provided by the ActiSleep monitor, a single-item self-report question was completed at each evaluation for comparison. The question, derived from the Sweat Index [63], was "How many days per week are you active enough where you begin to sweat?" Response options ranged from 0 to 7 days. Comparisons between baseline and follow-up were used to adjust modeling if activity levels increased or decreased significantly.

Adherence

Adherence was measured using the encrypted time and date stamped HR data obtained via the Tension Tamer app's built-in photoplethysmogram that were automatically relayed to our institution's server infrastructure. These data provided objective measures of adherence (ie, 1.0=2 sessions at full dose over a 24-hour period separated by >5 min; 0.5=completion of 1 session during the 24-hour period). Adherence was reported as a monthly percentage calculated as the average of daily scores (0, 0.5, and 1.0) across the number of days in the month.

Statistical Analyses

Demographic variables were analyzed using the analysis of variance for continuous data and chi-square tests for categorical data. For all analyses, statistical significance was set at P=.05. Analyses of longitudinal dose intervention effects from baseline to final 6-month evaluation for self-reported (PSQI) and actigraphy sleep indices were performed on the intention-to-treat (ITT) sample comprising all participants who completed at least four Tension Tamer sessions following completion of the baseline evaluation (n=57). A generalized linear mixed (GLM) model approach was used to account for the correlation of repeated measurements within patients [64,65]. The GLM models used the available measures at all time points (baseline and months 1, 3, and 6) and summarized the difference between groups at the final measurement (ie, month 6). A 2-level analysis strategy was utilized. Our primary analyses included the simple mixed-effects model containing the intervention dose group, the primary variable of interest, along with time (study visit), a dose-by-study visit interaction term as the fixed effect and adjusted for the baseline measure of the dependent variable with an unstructured covariance structure. The model was further adjusted for adherence.

The second level of GLM modeling included ethnicity as a fixed effect in addition to the previously included effects because of the statistically significant differences in most actigraphy-derived sleep indices observed at baseline. We estimated changes in sleep for each subject over the trial (preintervention and 1, 3, and 6 months) and the within-subject longitudinal trajectories (eg, slopes) and summarized the mean longitudinal trajectory within each dose group.

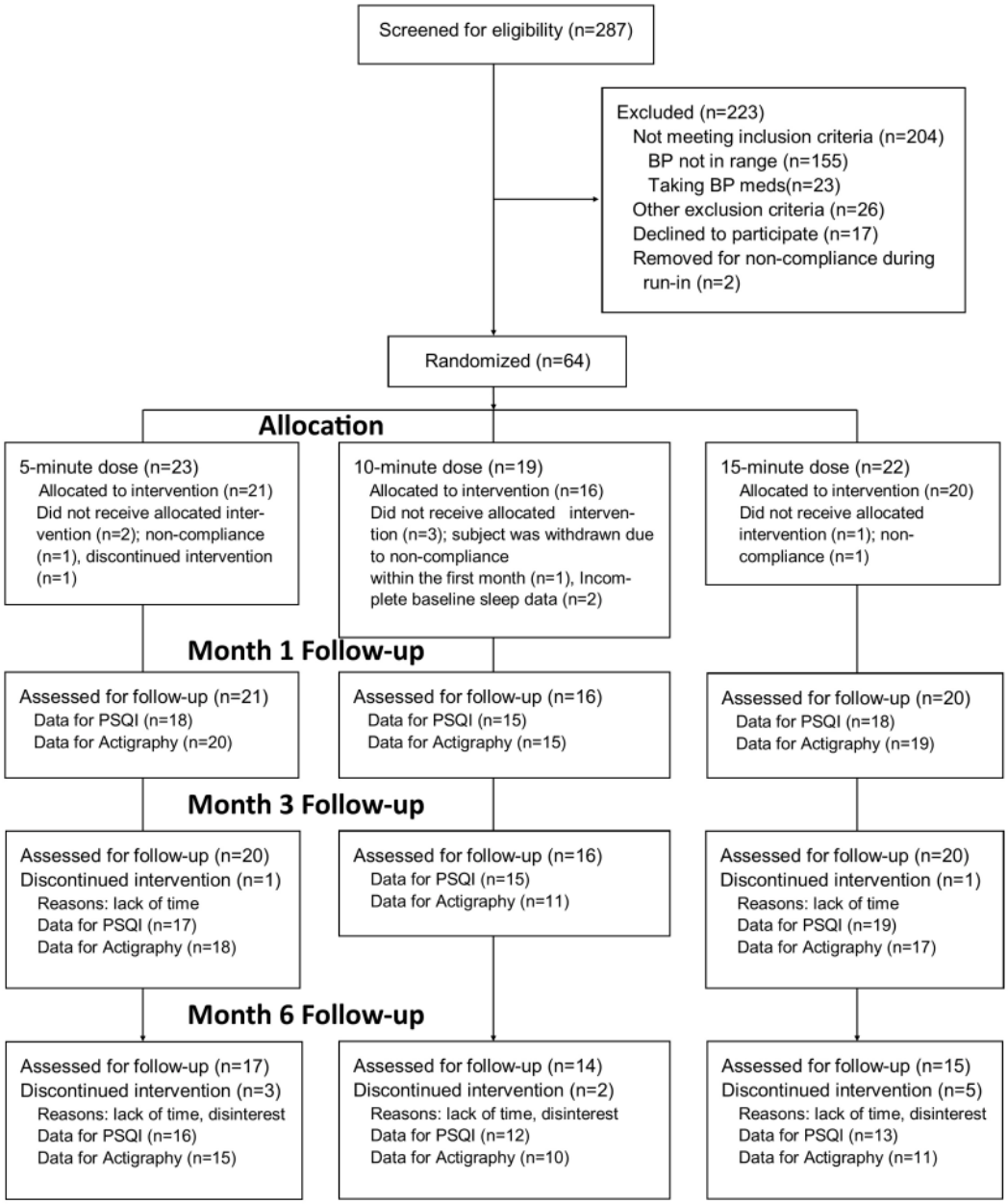
All sleep outcome variables were examined for normality using the Shapiro-Wilk test. With 18 subjects randomized to each intervention group, we had 80% power to detect at least a 0.96 SD unit effect size among the 3 groups (level of significance [alpha]=.05, 2-tailed) for sleep outcomes. SAS Statistical Software version 9.4 (SAS Institute Inc) and SPSS Statistics version 25.0 (IBM Corp) were used to perform the analyses.

Results

Overview

The study was approved in July 2014 (IRB#Pro00020894) and was performed between August, 2014 and October, 2016. A total of 287 potential candidates were screened (Figure 2). Overall, 204 patients were found to be ineligible. The primary reasons for exclusion included not meeting prehypertension criteria on all 3 screening visits (155/204, 76.0%), taking prescription medications that influenced BP (23/204, 11.2%), and failure to meet various other inclusion criteria (26/204, 12.7%), Of the 83 eligible participants, 64 (77%) consented to participate in the clinical trial. A total of 2 participants subsequently failed to complete the baseline evaluation monitoring in the natural environment. Furthermore, 5 others failed to complete at least 4 sessions during the first several weeks and dropped out (n=4) or were removed from the trial (n=3). These collective 7 individuals were evenly distributed across dosage conditions.

Figure 2. Recruitment and assessment flow diagram. PSQI: Pittsburgh Sleep Quality Index.



Baseline Characteristics

Overall, 47% (27/57) of the participants were female and 49% (28/57) were African American. As illustrated in Table 1, there were no significant baseline differences among the dosage groups on any descriptive characteristics (eg, education levels, income, BMI, physical activity, or stress; all values of P=.07 or greater). In addition, baseline sleep characteristics were not significantly different among dosage groups for any self-reported or objective variable (all P values were greater than .20).

AAs had higher BMI, reported less sweat-induced physical activity, and had lower education levels compared with NHWs (P=.04 to P=.001). Importantly, ethnicity differences were observed in 6 of the 7 baseline sleep quality indices (Table 2).

Table 1. Baseline demographic and sleep characteristics by dosage condition.

Characteristics	Overall (n=57)	Dose			*P* value[a]
		5 min (n=21)	10 min (n=16)	15 min (n=20)	
Age (years), mean (SD)	34.98 (12.79)	35.8 (11.10)	37.8 (13.9)	31.85 (13.48)	.36
Gender (female), n (%)	27 (47)	11 (52)	8 (50)	8 (40)	.71
Ethnicity (African American), n (%)	28 (49)	12 (67)	8 (50)	8 (40)	.55
BMI (kg/m^2), mean (SD)	27.63 (5.7)	28.5 (7.00)	26.4 (4.59)	27.7 (4.9)	.57
Physical activity (days/week), mean (SD)	3.32 (216)	3.1 (2.08)	3.1 (2.13)	3.7 (2.1)	.65
Annual income (US$), n (%)					.28
<15,000	20 (35)	6 (29)	4 (25)	10 (50)	
15,000-29,999	20 (35)	7 (33)	7 (44)	6 (30)	
30,000-49,999	10 (18)	4 (19)	4 (25)	2 (10)	
50,000-74,999	2 (4)	0 (0)	0 (0)	2 (10)	
>75,000	1 (2)	1 (5)	0 (0)	0 (0)	
Not reported	4 (7)	3 (14)	1 (6)	0 (0)	
Education, n (%)					.92
High school	5 (9)	2 (10)	2 (13)	5 (1)	
Trade school	10 (18)	5 (24)	2 (13)	3 (15)	
College educated	32 (56)	10 (48)	9 (56)	13 (65)	
Not reported	10 (18)	4 (19)	3 (19)	3 (15)	
Actigraphy indices, mean (SD)					
Sleep efficiency (%)	86.97 (6.62)	87.74 (5.45)	88.06 (6.40)	85.32 (7.82)	.40
Sleep duration (min/night)	380.24 (61.96)	379.88 (65.44)	373.17 (76.08)	386.20 (46.92)	.84
Sleep fragmentation (units)	26.76 (8.24)	25.67 (8.56)	25.06 (6.86)	29.26 (8.69)	.26
Valid days (days)	6.80 (0.49)	6.95 (0.22)	6.73 (0.59)	6.68 (0.58)	.20
Pittsburgh Sleep Quality Index domains, mean (SD)					
Sleep quality	2.3 (1.4)	1.9 (1.4)	2.3 (1.5)	2.6 (1.4)	.23
Sleep efficiency	1.3 (.96)	1.2 (1.1)	1.5 (1.0)	1.2 (.8)	.50
Daily disturbances	1.9 (1.0)	1.9 (1.1)	2.0 (1.2)	2.1 (.9)	.76
Global score	5.6 (2.6)	5.0 (2.8)	5.8 (2.7)	5.9 (2.4)	.44

[a]*P* values were determined by using the analysis of variance and the chi-square test (n=57).

Table 2. Baseline demographic and sleep characteristics by ethnicity.

Characteristics	Non-Hispanic White (n=29)	African American (n=28)	*P* value[a]
Age (years), mean (SD)	32.07 (10.84)	38.00 (14.10)	.08
Gender (female), n (%)	11 (38)	16 (57)	.15
BMI (kg/m^2), mean (SD)	25.47 (4.12)	29.87 (6.20)	.003
Physical activity (days/week), mean (SD)	3.86 (2.21)	2.74 (1.75)	.04
Annual income (US$), n (%)			.31
<15,000	45 (13)	7 (25)	
15,000-29,999	9 (31)	11 (39)	
30,000-49,999	5 (17)	5 (18)	
50,000-74,999	0 (0)	2 (7)	
>75,000	1 (3)	0 (0)	
Not reported	1 (3)	3 (11)	
Education, n (%)			.001
High school	0 (0)	5 (8)	
Trade school	1 (3)	9 (32)	
College educated	21 (72)	11 (39)	
Not reported	7 (24)	3 (11)	
Actigraphy indices, mean (SD)			
Sleep efficiency (%)	89.36 (4.71)	84.21 (7.48)	.003
Sleep duration (min/night)	413.25 (42.58)	341.00 (59.33)	.001
Sleep fragmentation (units)	24.34 (5.39)	29.57 (10.03)	.02
Pittsburgh Sleep Quality Index domains, mean (SD)			
Sleep quality	1.9 (1.4)	2.7 (1.3)	.03
Sleep efficiency	0.8 (0.7)	1.8 (0.9)	<.001
Daily disturbances	2.0 (1.1)	2.0 (1.0)	.89
Global score	4.7 (2.5)	6.5 (2.5)	.008

[a]*P* values were calculated using the analysis of variance and chi-square tests (n=57).

On the basis of the PSQI, AAs self-reported significantly poorer sleep efficiency (ie, shorter average sleep duration, higher percentage of time awake while trying to sleep) and poorer sleep quality (ie, using sleep medications, longer latency, poorer overall sleep quality); *P*<.001 and .03, respectively. As expected, their PSQI global sleep quality scores were higher (*P*=.008), with 50% (14/28) falling above the clinical cutoff level for sleep disturbance (PSQI global score >5) [50,58,61] compared with 13% (4/29) among NHWs. Importantly, all 3 actigraphy-derived indices supported the self-reported findings, with AAs exhibiting poorer sleep efficiency, greater fragmentation, and shorter average sleep duration than NHWs (*P*=.02 to *P*=.001).

Impact of Tension Tamer dosage on Sleep

Longitudinal GLM models evaluating the impact of Tension Tamer dosage conditions on sleep variables at 1, 3, and 6 months, controlling for baseline levels and adherence, did not reveal any statistically significant dose×time interaction terms in any of the PSQI-based sleep outcomes (all *P*=.08 or greater). Among the ActiGraph-derived variables, a significant dose effect was observed only for sleep fragmentation (*P*=.03). Collapsed across the 3 evaluation time points, the 5-min dose condition exhibited the lowest sleep fragmentation (23.4, SD 1.2) compared with both the 10-min (28.3, SD 1.5; *P*<.001) and 15-min (28.7, SD 1.3; *P*=.006) dose conditions.

Impact of Tension Tamer Dosage on Sleep Measures Within Ethnic Groups

GLM modeling was conducted with ethnicity as a covariate in the model (Table 3). Ethnicity was found to have a significant impact on the models' dosage results on 2 of the 3 actigraphic measures (fragmentation and efficiency; for both, *P*=.01) and on the PSQI sleep efficiency factor (*P*=.03).

Table 3. Effect of the dosage of Tension Tamer app on sleep measures within ethnic groups (adjusted means with SE).

Characteristics	Dose, 5 min	Dose, 10 min	Dose,15 min	Diff 5-10	*P* value	Diff 5-15	*P* value	Diff 10-15	*P* value
Sleep outcomes for non-Hispanic Whites[a]									
Actigraphy indices, mean (SE)									
Sleep efficiency (%)	91.4 (1.2)	87.2 (1.4)	89.7 (1.3)	4.2 (1.8)	.03	1.7 (1.8)	.35	−2.5 (1.9)	.19
Sleep fragment	21.9 (2.3)	26.1 (2.7)	25.5 (2.6)	−4.2 (1.8)	.25	−3.6 (3.5)	.32	0.6 (3.7)	.87
Sleep duration (min/night)	419.9 (12.8)	403.5 (14.7)	403.1 (14.0)	16.4 (19.5)	.41	16.8 (19.3)	.39	0.40 (19.8)	.98
PSQI[b], mean (SE)									
Sleep quality	0.7 (0.3)	1.5 (0.3)	2.2 (0.4)	−0.7 (0.5)	.02	−1.5 (0.5)	.006	−0.8 (0.5)	.13
Sleep efficiency	0.3 (0.2)	0.9 (0.2)	0.9 (0.2)	−0.6 (0.3)	.11	−0.7 (0.3)	.02	−0.1 (0.2)	.83
Daily disturbances	1.4 (02)	1.6 (0.3)	2.1 (0.3)	−0.2 (0.4)	.57	−0.7 (0.4)	.09	−0.5 (0.4)	.23
Global score	2.5 (0.4)	4.0 (0.4)	5.2 (0.4)	−1.4 (0.6)	.03	−2.7 (0.6)	<.001	−1.3 (0.6)	.04
Sleep outcomes for African Americans									
Actigraphy indices, mean (SE)									
Sleep efficiency (%)	85.9 (2.0)	82.9 (2.7)	83.0 (2.6)	3.0 (3.3)	.38	2.9 (3.2)	.38	−0.1 (3.8)	.99
Sleep fragment	24.5 (1.7)	30.9 (2.2)	37.4 (2.3)	−6.4 (2.7)	.03	−12.0 (3.1)	<.001	−6.7 (3.4)	.07
Sleep duration (min/night)	334.2 (9.9)	374.6 (12.6)	294.0 (13.3)	−40.5 (15.9)	.02	40.1 (17.0)	.03	80.6 (19.1)	<.001
PSQI, mean (SE)									
Sleep quality	2.2 (0.6)	2.2 (0.6)	1.7 (0.8)	−0.0 (0.9)	.99	0.4 (0.9)	.64	0.4 (0.9)	.63
Sleep efficiency	2.0 (0.3)	1.6 (0.3)	2.0 (0.3)	0.4 (0.4)	.35	0.04 (0.5)	.93	−0.4 (0.5)	.44
Daily disturbances	1.6 (0.4)	2.2 (0.4)	1.1 (0.5)	−0.6 (0.6)	.29	0.5 (0.6)	.64	1.1 (0.6)	.08
Global score	6.0 (1.0)	6.0 (1.0)	4.9 (1.2)	0.0 (1.5)	.99	1.1 (1.6)	.50	1.1 (1.5)	.50

[a]Differences determined by longitudinal generalized linear mixed modeling adjusted for visit, dose-by-visit interaction, adherence, and baseline outcome level (follow-up time points include 1, 3, and 6 months).

[b]PSQI: Pittsburgh Sleep Quality Index.

As a result of the above findings, an ethnicity×dose interaction term was added to the model to determine whether ethnicity modulated the effects of dosage level on sleep outcomes. Significant interaction effects were observed for 2 of the 3 actigraphic measures (efficiency *P*=.05; duration *P*=.03). The sleep measure was the dependent variable, whereas dosage was the primary independent variable adjusted for visit, dose-by-visit interaction term, baseline level of particular outcome measure, and adherence.

The NHWs exhibited a consistent pattern across both self-reported and ActiGraph-derived measures. The 5-min dose resulted in the greatest improvements in sleep measures compared with the other 2 dosage conditions. On the basis of the PSQI data, participants in the 5-min condition reported significantly better values for the sleep efficiency factor compared with both the 10-min and 15-min conditions (for both, *P*=.02). Similarly, they also reported better scores for the sleep quality factor compared with the 10-min and 15-min dose conditions, with the difference for the 15-min condition being statistically significant (*P*=.006). Finally, the 5-min condition reported significantly better global sleep quality compared with both the 10-min and 15-min conditions (*P*=.03 and <.001, respectively).

The ActiGraph findings supported these patterns among NHWs, with the 5-min group showing higher sleep efficiency compared with the 10-min (*P*=.03) and 15-min groups; however, the latter difference did not reach statistical significance. Similarly, the 5-min condition exhibited lower fragmentation and longer sleep duration compared with the other 2 conditions, but the differences were not statistically significant.

Among AAs, the actigraphy findings indicated that the 5-min condition exhibited significantly lower sleep fragmentation than both the 10-min and 15-min conditions (*P*=.03 and *P*<.001, respectively). The 10-min condition showed the highest sleep duration, which was significantly longer than both the 5-min and 15-min conditions (*P*=.02 and *P*<.001, respectively). The 5-min condition also exhibited significantly longer sleep duration compared with the 15-min condition (*P*=.03).

With regard to the self-reported results among AAs, none of the PSQI factor scores were significantly and differentially impacted by the Tension Tamer dosage, except for PSQI daily disturbances. The 15-min group showed a trend of lower daily

disturbances compared with the 10-min group (*P*=.08). Furthermore, the overall pattern of PSQI factor scores did not emulate the actigraphy findings. For example, the 15-min dose condition, which had the shortest actigraphy-derived sleep duration and greatest fragmentation during sleep, had the best sleep quality factor, global sleep quality scores, and daily disturbance factor score.

Effect of Differences in Tension Tamer Dosages on Sleep Measures Between Ethnic Groups

Further modeling was carried out to examine potential ethnicity differences in response to the 3 dosage levels given in Table 4. Adjusted means and SE by ethnicity within each dosage level at month 6 were obtained from longitudinal GLM modeling. The model included sleep measure as the dependent variable and dose as the primary independent variable of interest adjusted for visit, ethnicity, dose-by-visit and dose-by-visit-by-ethnicity interaction terms, baseline value of sleep measure, and adherence.

Per the PSQI outcomes, NHWs reported a consistent pattern of greater improvement than AAs for each PSQI factor (sleep quality, sleep efficiency, and daily disturbances) and global sleep quality for the 5-min and 10-min dose conditions. However, the only statistically significant differences observed were for the 5-min dose condition, in which NHWs reported greater sleep efficiency (*P*=.002) and better global sleep quality (*P*=.05) than AAs. The 15-min dose condition revealed a different pattern in which AAs reported better scores in sleep quality and daily disturbance factors and lower global sleep quality scores than NHWs; however, none of these differences reached statistical significance (all values of *P*=.18 or greater).

The ActiGraph results revealed a pattern in which NHWs exhibited comparable or better sleep efficiency and fragmentation than AAs across all 3 doses. They also exhibited longer sleep durations at the 5-min and 15-min dose conditions. However, statistically significant differences were observed in the 15-min dose condition only. NHWs exhibited better efficiency (*P*=.02), lower fragmentation (*P*=.02), and longer sleep duration (*P*=.002) than AAs.

Table 4. Effect of ethnicity differences and the dosage of Tension Tamer app on sleep measures at 6-month follow-upa. Values represent the adjusted means and SE.

Characteristics	Non-Hispanic Whites	African Americans	*P* value
Sleep outcomes for 5-min dose[a]			
Actigraphy, mean (SE)			
Sleep efficiency (%)	91.4 (1.2)	85.9 (2.0)	.18
Sleep fragment	21.9 (2.3)	24.5 (1.7)	.93
Sleep duration (min/night)	419.9 (12.8)	334.2 (9.9)	.13
PSQI [b] **, mean (SE)**			
Sleep quality	0.7 (0.3)	2.2 (0.6)	.14
Sleep efficiency	0.3 (0.2)	2.0 (0.3)	<.001
Daily disturbances	1.4 (02)	1.6 (0.4)	.56
Global score	2.5 (0.4)	6.0 (1.0)	.05
Sleep outcomes for 10-min dose			
Actigraphy, mean (SE)			
Sleep efficiency (%)	87.2 (1.4)	82.9 (2.7)	.66
Sleep fragment	26.1 (2.7)	30.9 (2.2)	.32
Sleep duration (min/night)	403.50 (14.7)	374.6 (12.6)	.28
PSQI, mean (SE)			
Sleep quality	1.5 (0.3)	2.2 (0.6)	.57
Sleep efficiency	0.9 (0.2)	1.6 (0.3)	.60
Daily disturbances	1.6 (0.3)	2.2 (0.4)	.13
Global score	4.0 (0.4)	6.0 (1.0)	.32
Sleep outcomes for 15-min dose			
Actigraphy, mean (SE)			
Sleep efficiency (%)	89.7 (1.3)	83.0 (2.6)	.02
Sleep fragment	25.5 (2.6)	37.4 (2.3)	.02
Sleep duration (min/night)	403.1 (14.0)	294.0 (13.3)	.002
PSQI, mean (SE)			
Sleep quality	2.2 (0.4)	1.7 (0.8)	.48
Sleep efficiency	0.9 (0.2)	2.0 (0.3)	.12
Daily disturbances	2.1 (0.3)	1.1 (0.5)	.18
Global score	5.2 (0.4)	4.9 (1.2)	.54

[a]Differences determined by longitudinal generalized linear mixed modeling adjusted for visit, dose-by-visit interaction, adherence, and baseline outcome level.

[b]PSQI: Pittsburgh Sleep Quality Index.

Discussion

Principal Findings

This is the first study to use a smartphone app to examine the effect of varying doses of BAM on self-reported and objective indices of sleep. Baseline measurements corroborated the literature with regard to adult ethnicity differences in sleep quality [27-31,35,57-59,66]. Similar to most research that has utilized self-reported methodology, we used the well-validated PSQI survey [50,56] and a 3-factor model first reported by Cole et al [57] and replicated it across multiple patient populations and ethnic groups, including NHWs and AAs [59-61]. AAs, compared with NHWs, reported statistically significant poorer scores for sleep efficiency (ie, shorter average sleep duration, higher percentage time awake while trying to sleep) and sleep quality (ie, using sleep medications, longer latency, poorer general sleep quality) factors. On the basis of the clinical cutoff point for poor sleep quality (ie, PSQI global score of >5) [50,58,61], AAs had a significantly higher prevalence of poor sleep quality compared with NHWs.

The 24-hour monitoring with the ActiGraph wrist monitor for 7 consecutive days at each evaluation point provided objective indices of sleep. Importantly, these data corroborated the baseline self-reported findings in that AAs exhibited shorter sleep duration, more fragmented sleep, and lower sleep efficiency.

Longitudinal modeling analyses indicated that the BAM program, delivered and monitored using the Tension Tamer smartphone app, had a beneficial impact on sleep, which was modulated by ethnicity. On the basis of actigraphy monitoring, among AAs, the 5-min dose resulted in significantly lower fragmentation compared with both the 10-min and 15-min dose conditions. The 5-min dose group also had significantly longer sleep duration than the 15-min dose group; however, the 10-min dose resulted in the longest sleep duration compared with both the 5-min and 15-min doses. Previously, we compared the changes in SBP among dosage conditions and observed that the 15-min condition had significantly lower adherence rates compared with the other 2 conditions [23]. By 6 months, the ITT analysis using 75% adherence criterion found that the 5-min condition had higher percentages (57%) meeting this criterion compared with 37% for the 10-min condition and 14% for the 15-min condition.

Among the AAs, no significant differential impact by dosage condition was observed on self-reported sleep. Furthermore, the general pattern of PSQI-derived sleep dimensions did not corroborate the actigraphy findings. This discrepancy is akin to the findings that AAs exhibit lower correlations between self-reported sleep duration and both actigraphy- and polysomnography-derived sleep durations (r range 0.15 to 0.29) compared with NHWs (r range 0.45 to 0.56) [33,67-69].

On the other hand, among the NHWs, a consistent pattern across both self-reported and ActiGraph-derived indices indicated that the 5-min dose resulted in the greatest improvements in sleep measures. On the basis of the PSQI, the 5-min condition reported significantly greater improvements in sleep efficiency compared with both the 10-min and 15-min conditions. Similarly, the 5-min condition reported better sleep quality compared with the other 2 conditions, with the difference for the 15-min condition being statistically significant. Finally, the 5-min dose condition reported significantly better PSQI global scores (summation of the original 7 sleep components) compared with both 10-min and 15-min dose conditions. The pattern of actigraphy findings supported the self-reported findings. The 5-min condition exhibited higher sleep efficiency scores compared with both 10-min and 15-min conditions, with the 10-min comparison being statistically significant. Similarly, the 5-min condition had lower fragmentation compared with the other 2 dosage conditions; however, the differences did not reach statistical significance.

Evaluations of potential ethnicity differences in response to the 3 Tension Tamer doses revealed a consistent pattern of significantly better responsivity among the NHWs. ActiGraph-derived measures indicated significantly greater improvements among NHWs in sleep efficiency, reduced fragmentation, and longer sleep duration compared with AAs at each dosage condition. The PSQI findings provided support to the actigraphy findings, with NHWs reporting a pattern of higher quality, efficiency, and better PSQI global sleep scores for both 5-min and 10-min dose conditions compared with AAs. Compared with NHWs, AAs in the 15-min dose condition reported better sleep quality and PSQI daily disturbance scores, with a trend toward better global sleep scores. A higher use of Tension Tamer in the 5-min group across the trial may have contributed to the generally greater beneficial influence on sleep quality. Anecdotally, the 5-min participants appeared to more often note at posttrial key informant interviews using Tension Tamer immediately before going to bed, which may have played a role as well.

As noted earlier, the findings have been mixed with regard to the impact of meditation on sleep [9,11,13,17,24]. The overall results of this study provide increasing support for the role of smartphone-enabled meditation and the benefits of multimethod-based sleep outcomes among individuals not originally recruited for sleep disturbance studies [70-72]. As pointed out in a review of 123 mind-body based interventions by Neuendorf et al [11], the use of meditation may be advantageous to improve sleep among individuals who do not have clinical levels of sleep disturbance, but the observed effect sizes are likely to be less. The mixed sleep results observed in earlier studies may be due to poor scientific rigor, lack of measurement of adherence, or dosage of mindfulness-based practices, especially in studies conducted as home-based programs [13]. These shortcomings are addressed in this report where dosage and adherence measures bolster the quality of our findings.

Use of cellular technology provides the possibility of disseminating efficacious mind-body intervention programs such as BAM to larger audiences, circumventing barriers such as travel, childcare, and other expense-related issues. In addition, the penetration of smartphone-assisted technology into lower income, rural, and urban areas enables these programs to be widely disseminated and available. This is increasingly important as 81% of adults own a smartphone as of June 2019 [73]. Our exit surveys suggest that Tension Tamer sessions conducted before bedtime may aid individuals in going to sleep faster. Thus, engagement with Tension Tamer as part of the sleep preparation ritual may be beneficial for future studies. Tension Tamer may potentially have a larger effect in improving sleep if used as a direct intervention or in combination with other sleep hygiene practices on those exhibiting sleep complaints. Further iterative, user-centered design is needed with NHWs and AAs with sleep disturbances to understand their levels of sleep hygiene literacy and identify potential needs for culturally attuned tailoring of a refined Tension Tamer app that would include sleep hygiene education, video testimonies by NHWs and AAs who had experienced sleep issues and benefited from Tension Tamer, and video segments by such individuals showing how to use a refined Tension Tamer. This aligns with recent recommendations by Johnson et al [68], who emphasized the need for culturally sensitive efforts to reduce racial and ethnic sleep disparities. The study emphatically reports the disparities in AAs among the sleep dimensions of duration, quality, and sleepiness. Specifically, sleep programs involving ethnic minorities will likely need to include sleep

hygiene education to help eradicate false beliefs on managing issues such as long sleep latency (eg, to fall asleep, just watch TV in bed or drink alcohol).

Limitations

A limitation of this study was that the Tension Tamer app was not designed to specifically target individuals with sleep issues but was rather for those with prehypertension. As a result, we did not include eligibility criteria related to sleep problems, insomnia, or collect sleep medication usage. Therefore, bias toward the null hypothesis may be an explanation for some of the findings. In future studies, inclusion of targeted intervention materials in addressing sleep-related barriers in prehypertensive patients may have a greater effect. Such tactics could include increasing basic sleep hygiene awareness or using the Tension Tamer app in combination with cognitive behavioral therapy [68]. Furthermore, 2 other limitations, the lack of a control arm and the relatively small sample size in each dosage arm, limit the generalizability of the findings. Comparisons between ethnic groups also found that the AA sample was older, had higher BMIs, and lower levels of physical activity, which could have contributed to poorer sleep quality results.

This study was also not designed to examine the daily timing of BAM sessions around bedtime routines. A properly powered trial using a refined Tension Tamer app, including BAM as part of bedtime preparation rituals, is needed to understand how the timing and dosage of BAM or other meditation techniques may provide the most beneficial effect on sleep outcomes. We also did not investigate any related mechanistic evaluations such as electromyography or other common sleep tests such as polysomnography [74], which could impart important information on how BAM affects sleep using causal pathways.

Conclusions

To our knowledge, this is the only study that has examined the influence of varying doses of meditation delivered via a smartphone app on multimethod-derived indices of sleep (ie, self-report and actigraphy) in a multiethnic population not diagnosed with sleep disturbance issues. A high level of methodological rigor was incorporated, including random assignment to dosage condition, automated feedback of dosage adherence, and postsession HR graphs to provide reinforcement for engaging with the Tension Tamer app. Ethnicity was found to play a modulating role on the effects of Tension Tamer. Formative evaluation adapted from the development of Tension Tamer [75] can target ethnicity-related sleep issues utilizing a theory-guided, user-centered iterative design. This could aid in using the Tension Tamer app to facilitate sleep hygiene education, dispel myths, incorporate sleep ritual behaviors, and further address ethnic sleep disparities, as suggested by Johnson et al [68]. Future studies should further investigate the potential for ethnicity-induced differences in self-reported and objectively derived sleep measures using varying doses of BAM or other types of meditation or cognitive behavioral sleep programs, especially among those with verified clinical levels of sleep disturbance. Issues to address include levels of adherence to the regimen, potential influence of timing of daily engagement, and how it may differentially affect barriers to sleep. Finally, future studies will benefit from identifying the underlying causal mechanisms for BAM's beneficial impact on sleep (eg, alterations in sleep stages, HR variability, hypothalamic-pituitary-adrenal axis, sympathetic nervous system activation, rumination, worry, intrusive thoughts) [76].

Acknowledgments

This publication was supported by funding from the National Institutes of Health (NIH) grants R01 HL114957 and K23 DA038257 and by the South Carolina Clinical and Translational Research Institute, with an academic home at Medical University of South Carolina, through NIH grant number UL1 TR000062.

Conflicts of Interest

None declared.

References

1. Elder CR, Gullion CM, Funk KL, Debar LL, Lindberg NM, Stevens VJ. Impact of sleep, screen time, depression and stress on weight change in the intensive weight loss phase of the LIFE study. Int J Obes (Lond) 2012 Jan;36(1):86-92 [FREE Full text] [doi: 10.1038/ijo.2011.60] [Medline: 21448129]
2. Consensus Conference Panel, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, et al. Joint consensus statement of the American academy of sleep medicine and sleep research society on the recommended amount of sleep for a healthy adult: methodology and discussion. J Clin Sleep Med 2015 Aug 15;11(8):931-952 [FREE Full text] [doi: 10.5664/jcsm.4950] [Medline: 26235159]
3. Grandner MA, Chakravorty S, Perlis ML, Oliver L, Gurubhagavatula I. Habitual sleep duration associated with self-reported and objectively determined cardiometabolic risk factors. Sleep Med 2014 Jan;15(1):42-50 [FREE Full text] [doi: 10.1016/j.sleep.2013.09.012] [Medline: 24333222]
4. Morselli LL, Guyon A, Spiegel K. Sleep and metabolic function. Pflugers Arch 2012 Jan;463(1):139-160 [FREE Full text] [doi: 10.1007/s00424-011-1053-z] [Medline: 22101912]

5. Zizi F, Pandey A, Murrray-Bachmann R, Vincent M, McFarlane S, Ogedegbe G, et al. Race/ethnicity, sleep duration, and diabetes mellitus: analysis of the national health interview survey. Am J Med 2012 Feb;125(2):162-167 [FREE Full text] [doi: 10.1016/j.amjmed.2011.08.020] [Medline: 22269619]
6. Buxton OM, Marcelli E. Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States. Soc Sci Med 2010 Sep;71(5):1027-1036. [doi: 10.1016/j.socscimed.2010.05.041] [Medline: 20621406]
7. Cvengros JA, Crawford MR, Manber R, Ong JC. The relationship between beliefs about sleep and adherence to behavioral treatment combined with meditation for insomnia. Behav Sleep Med 2015;13(1):52-63 [FREE Full text] [doi: 10.1080/15402002.2013.838767] [Medline: 24354360]
8. Drake CL, Pillai V, Roth T. Stress and sleep reactivity: a prospective investigation of the stress-diathesis model of insomnia. Sleep 2014 Aug 01;37(8):1295-1304 [FREE Full text] [doi: 10.5665/sleep.3916] [Medline: 25083009]
9. Kim SM, Park JM, Seo H. Effects of mindfulness-based stress reduction for adults with sleep disturbance: a protocol for an update of a systematic review and meta-analysis. Syst Rev 2016 Apr 2;5:51 [FREE Full text] [doi: 10.1186/s13643-016-0228-2] [Medline: 27039290]
10. Nagendra RP, Maruthai N, Kutty BM. Meditation and its regulatory role on sleep. Front Neurol 2012;3:54 [FREE Full text] [doi: 10.3389/fneur.2012.00054] [Medline: 22529834]
11. Neuendorf R, Wahbeh H, Chamine I, Yu J, Hutchison K, Oken BS. The effects of mind-body interventions on sleep quality: a systematic review. Evid Based Complement Alternat Med 2015;2015:902708 [FREE Full text] [doi: 10.1155/2015/902708] [Medline: 26161128]
12. Pillai V, Steenburg LA, Ciesla JA, Roth T, Drake CL. A seven day actigraphy-based study of rumination and sleep disturbance among young adults with depressive symptoms. J Psychosom Res 2014 Jul;77(1):70-75. [doi: 10.1016/j.jpsychores.2014.05.004] [Medline: 24913345]
13. Winbush NY, Gross CR, Kreitzer MJ. The effects of mindfulness-based stress reduction on sleep disturbance: a systematic review. Explore (NY) 2007;3(6):585-591. [doi: 10.1016/j.explore.2007.08.003] [Medline: 18005910]
14. Priya G, Kalra S. Mind-body interactions and mindfulness meditation in diabetes. Eur Endocrinol 2018 Apr;14(1):35-41 [FREE Full text] [doi: 10.17925/EE.2018.14.1.35] [Medline: 29922350]
15. Park S, Han KS. Blood pressure response to meditation and yoga: a systematic review and meta-analysis. J Altern Complement Med 2017 Sep;23(9):685-695. [doi: 10.1089/acm.2016.0234] [Medline: 28384004]
16. Carrière K, Khoury B, Günak MM, Knäuper B. Mindfulness-based interventions for weight loss: a systematic review and meta-analysis. Obes Rev 2018 Feb;19(2):164-177. [doi: 10.1111/obr.12623] [Medline: 29076610]
17. Goyal M, Singh S, Sibinga EM, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. JAMA Intern Med 2014 Mar;174(3):357-368 [FREE Full text] [doi: 10.1001/jamainternmed.2013.13018] [Medline: 24395196]
18. Nidich SI, Rainforth MV, Haaga DA, Hagelin J, Salerno JW, Travis F, et al. A randomized controlled trial on effects of the transcendental meditation program on blood pressure, psychological distress, and coping in young adults. Am J Hypertens 2009 Dec;22(12):1326-1331 [FREE Full text] [doi: 10.1038/ajh.2009.184] [Medline: 19798037]
19. Schneider RH, Staggers F, Alxander CN, Sheppard W, Rainforth M, Kondwani K, et al. A randomised controlled trial of stress reduction for hypertension in older African Americans. Hypertension 1995 Nov;26(5):820-827. [doi: 10.1161/01.hyp.26.5.820] [Medline: 7591024]
20. Anderson JW, Liu C, Kryscio RJ. Blood pressure response to transcendental meditation: a meta-analysis. Am J Hypertens 2008 Mar;21(3):310-316. [doi: 10.1038/ajh.2007.65] [Medline: 18311126]
21. Wright LB, Gregoski MJ, Tingen MS, Barnes VA, Treiber FA. Impact of stress reduction interventions on hostility and ambulatory systolic blood pressure in African American adolescents. J Black Psychol 2011 May;37(2):210-233 [FREE Full text] [doi: 10.1177/0095798410380203] [Medline: 22485058]
22. Dickinson H, Campbell F, Beyer F, Nicolson D, Cook J, Ford G, et al. Relaxation therapies for the management of primary hypertension in adults: a Cochrane review. J Hum Hypertens 2008 Dec;22(12):809-820. [doi: 10.1038/jhh.2008.65] [Medline: 18548088]
23. Adams ZW, Sieverdes JC, Brunner-Jackson B, Mueller M, Chandler J, Diaz V, et al. Meditation smartphone application effects on prehypertensive adults' blood pressure: dose-response feasibility trial. Health Psychol 2018 Sep;37(9):850-860 [FREE Full text] [doi: 10.1037/hea0000584] [Medline: 30010353]
24. Gong H, Ni C, Liu Y, Zhang Y, Su W, Lian Y, et al. Mindfulness meditation for insomnia: a meta-analysis of randomized controlled trials. J Psychosom Res 2016 Oct;89:1-6. [doi: 10.1016/j.jpsychores.2016.07.016] [Medline: 27663102]
25. Martín DB, de la Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mhealth: systematic search in app stores. JMIR Mhealth Uhealth 2018 May 9;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
26. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 3;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]

27. Cunningham TJ, Ford ES, Chapman DP, Liu Y, Croft JB. Independent and joint associations of race/ethnicity and educational attainment with sleep-related symptoms in a population-based US sample. Prev Med 2015 Aug;77:99-105 [FREE Full text] [doi: 10.1016/j.ypmed.2015.05.008] [Medline: 26004167]
28. Hale L, Do DP. Racial differences in self-reports of sleep duration in a population-based study. Sleep 2007 Sep;30(9):1096-1103 [FREE Full text] [doi: 10.1093/sleep/30.9.1096] [Medline: 17910381]
29. Jean-Louis G, Magai CM, Cohen CI, Zizi F, von Gizycki H, DiPalma J, et al. Ethnic differences in self-reported sleep problems in older adults. Sleep 2001 Dec 15;24(8):926-933. [doi: 10.1093/sleep/24.8.926] [Medline: 11766163]
30. Patel NP, Grandner MA, Xie D, Branas CC, Gooneratne N. 'Sleep disparity' in the population: poor sleep quality is strongly associated with poverty and ethnicity. BMC Public Health 2010 Aug 11;10:475 [FREE Full text] [doi: 10.1186/1471-2458-10-475] [Medline: 20701789]
31. Williams NJ, Grandner MA, Wallace DM, Cuffee Y, Airhihenbuwa C, Okuyemi K, et al. Social and behavioral predictors of insufficient sleep among African Americans and Caucasians. Sleep Med 2016 Feb;18:103-107 [FREE Full text] [doi: 10.1016/j.sleep.2015.02.533] [Medline: 26514614]
32. Carnethon MR, De Chavez PJ, Zee PC, Kim KA, Liu K, Goldberger JJ, et al. Disparities in sleep characteristics by race/ethnicity in a population-based sample: Chicago area sleep study. Sleep Med 2016 Feb;18:50-55 [FREE Full text] [doi: 10.1016/j.sleep.2015.07.005] [Medline: 26459680]
33. Chen X, Wang R, Zee P, Lutsey PL, Javaheri S, Alcántara C, et al. Racial/ethnic differences in sleep disturbances: the multi-ethnic study of atherosclerosis (MESA). Sleep 2015 Jun 1;38(6):877-888 [FREE Full text] [doi: 10.5665/sleep.4732] [Medline: 25409106]
34. Matthews KA, Hall MH, Lee L, Kravitz HM, Chang Y, Appelhans BM, et al. Racial/ethnic disparities in women's sleep duration, continuity, and quality, and their statistical mediators: study of women's health across the nation. Sleep 2019 May 1;42(5):- [FREE Full text] [doi: 10.1093/sleep/zsz042] [Medline: 30778560]
35. Ruiter ME, Decoster J, Jacobs L, Lichstein KL. Normal sleep in African-Americans and Caucasian-Americans: a meta-analysis. Sleep Med 2011 Mar;12(3):209-214. [doi: 10.1016/j.sleep.2010.12.010] [Medline: 21317037]
36. Barnes VA, Gregoski MJ, Tingen MS, Treiber FA. Influences of family environment and meditation efficacy on hemodynamic function among African American adolescents. J Complement Integr Med 2010 Jul;7(1):- [FREE Full text] [doi: 10.2202/1553-3840.1326] [Medline: 22328869]
37. Barnes VA, Pendergrast RA, Harshfield GA, Treiber FA. Impact of breathing awareness meditation on ambulatory blood pressure and sodium handling in prehypertensive African American adolescents. Ethn Dis 2008;18(1):1-5 [FREE Full text] [Medline: 18447091]
38. Gregoski MJ, Vertegel A, Shaporev A, Treiber FA. Tension Tamer: delivering meditation with objective heart rate acquisition for adherence monitoring using a smart phone platform. J Altern Complement Med 2013 Jan;19(1):17-19 [FREE Full text] [doi: 10.1089/acm.2011.0772] [Medline: 22967280]
39. Ospina MB, Bond K, Karkhaneh M, Tjosvold L, Vandermeer B, Liang Y, et al. Meditation practices for health: state of the research. Evid Rep Technol Assess (Full Rep) 2007 Jun(155):1-263. [Medline: 17764203]
40. Rainforth MV, Schneider RH, Nidich SI, Gaylord-King C, Salerno JW, Anderson JW. Stress reduction programs in patients with elevated blood pressure: a systematic review and meta-analysis. Curr Hypertens Rep 2007 Dec;9(6):520-528 [FREE Full text] [doi: 10.1007/s11906-007-0094-3] [Medline: 18350109]
41. Anderson DE, McNeely JD, Windham BG. Regular slow-breathing exercise effects on blood pressure and breathing patterns at rest. J Hum Hypertens 2010 Dec;24(12):807-813. [doi: 10.1038/jhh.2010.18] [Medline: 20200548]
42. Kumar A, Goel H, Nadar SK. Short sleep duration and the risk of hypertension: snoozing away high blood pressure? J Hum Hypertens 2019 Mar;33(3):174-176. [doi: 10.1038/s41371-019-0177-z] [Medline: 30778131]
43. Han B, Chen WZ, Li YC, Chen J, Zeng ZQ. Sleep and hypertension. Sleep Breath 2020 Mar;24(1):351-356 [FREE Full text] [doi: 10.1007/s11325-019-01907-2] [Medline: 31402441]
44. Wang L, Hu Y, Wang X, Yang S, Chen W, Zeng Z. The association between sleep duration and hypertension: a meta and study sequential analysis. J Hum Hypertens 2020 Jun 25:- epub ahead of print. [doi: 10.1038/s41371-020-0372-y] [Medline: 32587332]
45. Gregoski MJ, Mueller M, Vertegel A, Shaporev A, Jackson BB, Frenzel RM, et al. Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. Int J Telemed Appl 2012;2012:696324 [FREE Full text] [doi: 10.1155/2012/696324] [Medline: 22272197]
46. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
47. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
48. Deci E, Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior. New York, USA: Plenum; 1985.
49. Benyon D, Turner P, Trurner S. Designing Interactive Systems: People, Activities, Contexts, Technologies. New York, USA: Addison-Wesley; 2005.

50. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
51. Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. Expert Rev Med Devices 2016 May;13(5):497-506. [doi: 10.1586/17434440.2016.1171708] [Medline: 27043070]
52. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Nov 1;36(11):1747-1755 [FREE Full text] [doi: 10.5665/sleep.3142] [Medline: 24179309]
53. Full KM, Kerr J, Grandner MA, Malhotra A, Moran K, Godoble S, et al. Validation of a physical activity accelerometer device worn on the hip and wrist against polysomnography. Sleep Health 2018 Apr;4(2):209-216 [FREE Full text] [doi: 10.1016/j.sleh.2017.12.007] [Medline: 29555136]
54. Quante M, Kaplan ER, Cailler M, Rueschman M, Wang R, Weng J, et al. Actigraphy-based sleep estimation in adolescents and adults: a comparison with polysomnography using two scoring algorithms. Nat Sci Sleep 2018;10:13-20 [FREE Full text] [doi: 10.2147/NSS.S151085] [Medline: 29403321]
55. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [doi: 10.1093/sleep/15.5.461] [Medline: 1455130]
56. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh sleep quality index. J Psychosom Res 1998 Jul;45(1):5-13. [doi: 10.1016/s0022-3999(97)00298-5] [Medline: 9720850]
57. Cole JC, Motivala SJ, Buysse DJ, Oxman MN, Levin MJ, Irwin MR. Validation of a 3-factor scoring model for the Pittsburgh sleep quality index in older adults. Sleep 2006 Jan;29(1):112-116. [doi: 10.1093/sleep/29.1.112] [Medline: 16453989]
58. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. Sleep Med Rev 2016 Feb;25:52-73. [doi: 10.1016/j.smrv.2015.01.009] [Medline: 26163057]
59. Tomfohr LM, Schweizer CA, Dimsdale JE, Loredo JS. Psychometric characteristics of the Pittsburgh sleep quality index in English speaking non-Hispanic whites and English and Spanish speaking Hispanics of Mexican descent. J Clin Sleep Med 2013 Jan 15;9(1):61-66 [FREE Full text] [doi: 10.5664/jcsm.2342] [Medline: 23319906]
60. Burkhalter H, Sereika S, Engberg S, Wirz-Justice A, Steiger J, De GS. Structure validity of the Pittsburgh sleep quality index in renal transplant recipients: a confirmatory factor analysis. Sleep Biol Rhythms 2018;8(4):274-281 [FREE Full text] [doi: 10.1111/j.1479-8425.2010.00473.x]
61. Otte JL, Rand KL, Landis CA, Paudel ML, Newton KM, Woods N, et al. Confirmatory factor analysis of the Pittsburgh sleep quality index in women with hot flashes. Menopause 2015 Nov;22(11):1190-1196 [FREE Full text] [doi: 10.1097/GME.0000000000000459] [Medline: 25944520]
62. Skouteris H, Wertheim EH, Germano C, Paxton SJ, Milgrom J. Assessing sleep during pregnancy: a study across two time points examining the Pittsburgh Sleep Quality Index and associations with depressive symptoms. Womens Health Issues 2009;19(1):45-51. [doi: 10.1016/j.whi.2008.10.004] [Medline: 19111787]
63. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item physical activity measure. Br J Sports Med 2011 Mar;45(3):203-208 [FREE Full text] [doi: 10.1136/bjsm.2009.068395] [Medline: 20484314]
64. Hanley JA, Negassa A, Edwardes MD, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol 2003 Feb 15;157(4):364-375. [doi: 10.1093/aje/kwf215] [Medline: 12578807]
65. Hardin J, Hilbe J. Generalized Estimating Equations. John Wiley & Sons Inc: Chapman and Hall/CRC Press; 2013.
66. Grandner MA, Petrov ME, Rattanaumpawan P, Jackson N, Platt A, Patel NP. Sleep symptoms, race/ethnicity, and socioeconomic position. J Clin Sleep Med 2013 Sep 15;9(9):897-905; 905A [FREE Full text] [doi: 10.5664/jcsm.2990] [Medline: 23997702]
67. Jackson CL, Patel SR, Jackson WB, Lutsey PL, Redline S. Agreement between self-reported and objectively measured sleep duration among white, black, Hispanic, and Chinese adults in the United States: multi-ethnic study of atherosclerosis. Sleep 2018 Jun 1;41(6):- [FREE Full text] [doi: 10.1093/sleep/zsy057] [Medline: 29701831]
68. Johnson DA, Jackson CL, Williams NJ, Alcántara C. Are sleep patterns influenced by race/ethnicity - a marker of relative advantage or disadvantage? Evidence to date. Nat Sci Sleep 2019;11:79-95 [FREE Full text] [doi: 10.2147/NSS.S169312] [Medline: 31440109]
69. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology 2008 Nov;19(6):838-845 [FREE Full text] [doi: 10.1097/EDE.0b013e318187a7b0] [Medline: 18854708]
70. Black DS, O'Reilly GA, Olmstead R, Breen EC, Irwin MR. Mindfulness meditation and improvement in sleep quality and daytime impairment among older adults with sleep disturbances: a randomized clinical trial. JAMA Intern Med 2015 Apr;175(4):494-501 [FREE Full text] [doi: 10.1001/jamainternmed.2014.8081] [Medline: 25686304]
71. Brand S, Holsboer-Trachsler E, Naranjo JR, Schmidt S. Influence of mindfulness practice on cortisol and sleep in long-term and short-term meditators. Neuropsychobiology 2012;65(3):109-118. [doi: 10.1159/000330362] [Medline: 22377965]
72. Ong JC, Manber R, Segal Z, Xia Y, Shapiro S, Wyatt JK. A randomized controlled trial of mindfulness meditation for chronic insomnia. Sleep 2014 Sep 1;37(9):1553-1563 [FREE Full text] [doi: 10.5665/sleep.4010] [Medline: 25142566]

73. Mobile Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-09-07]
74. Patra S, Telles S. Positive impact of cyclic mPositive impact of cyclic meditation on subsequent sleepeditation on subsequent sleep. Med Sci Monit 2009 Jul;15(7):CR375-CR381. [Medline: 19564829]
75. Sieverdes JC, Adams ZW, Nemeth L, Brunner-Jackson B, Mueller M, Anderson A, et al. Formative evaluation on cultural tailoring breathing awareness meditation smartphone apps to reduce stress and blood pressure. Mhealth 2017;3:44 [FREE Full text] [doi: 10.21037/mhealth.2017.09.04] [Medline: 29184896]
76. Sieverdes JC, Mueller M, Gregoski MJ, Brunner-Jackson B, McQuade L, Matthews C, et al. Effects of Hatha yoga on blood pressure, salivary α-amylase, and cortisol function among normotensive and prehypertensive youth. J Altern Complement Med 2014 Apr;20(4):241-250 [FREE Full text] [doi: 10.1089/acm.2013.0139] [Medline: 24620850]

Abbreviations

AA: African Americans
BAM: breathing awareness meditation
BP: blood pressure
GLM: generalized linear mixed
HIPAA: Health Insurance Portability and Accountability Act
HR: heart rate
ITT: intention-to-treat
LED: light-emitting diode
NHW: non-Hispanic Whites
NIH: National Institutes of Health
PACT: People, Activity, Context, and Technology
PSQI: Pittsburgh Sleep Quality Index
PTSD: post-traumatic stress disorder
RCT: randomized controlled trial
SBP: systolic blood pressure
TM: transcendental meditation

Edited by G Eysenbach; submitted 20.05.20; peer-reviewed by V Rocío; comments to author 04.07.20; revised version received 22.07.20; accepted 26.07.20; published 06.10.20.

Please cite as:
Sieverdes JC, Treiber FA, Kline CE, Mueller M, Brunner-Jackson B, Sox L, Cain M, Swem M, Diaz V, Chandler J
Ethnicity Differences in Sleep Changes Among Prehypertensive Adults Using a Smartphone Meditation App: Dose-Response Trial
JMIR Form Res 2020;4(10):e20501
URL: https://formative.jmir.org/2020/10/e20501
doi:10.2196/20501
PMID:33021484

Original Paper

Using ADAPT-ITT to Modify a Telephone-Based HIV Prevention Intervention for SMS Delivery: Formative Study

Teaniese Davis[1*], MPH, PhD; Ralph Joseph DiClemente[2*], MSc, PhD; Michael Prietula[3*], MPH, PhD

[1]Center for Research and Evaluation, Kaiser Permanente Georgia, Atlanta, GA, United States

[2]Department of Social & Behavioral Sciences, School of Global Public Health, New York University, New York, NY, United States

[3]Goizueta Business School & Hubert Department of Global Health, Emory University, Atlanta, GA, United States

[*]all authors contributed equally

Corresponding Author:
Michael Prietula, MPH, PhD
Goizueta Business School & Hubert Department of Global Health
Emory University
Goizueta Business School
1300 Clifton Road
Atlanta, GA, 30322
United States
Phone: 1 7709009034
Email: mj.prietula@emory.edu

Abstract

Background: African American adolescent females are disproportionately affected by sexually transmitted infections (STIs) and HIV. Given the elevated risk of STIs and HIV in African American women, there is an urgent need to identify innovative strategies to enhance the adoption and maintenance of STI and HIV preventive behaviors. Texting is a promising technology for creating preventive maintenance interventions (PMIs) that extend the efficacy of the original intervention. However, little guidance in public health literature is available for developing this type of application.

Objective: This paper describes a formative pilot study that incorporates user experience methods to design and test PMI texts for Afiya, an original evidence-based intervention (EBI) specifically designed for African American adolescent females. This study aims to describe the adaptation process of health educator–led phone calling to text-based communication.

Methods: The formative process followed the assessment, decision, adaptation, production, topical experts-integration, training, testing (ADAPT-ITT) framework for adapting EBIs and using them in a new setting, for a new target population or a modified intervention strategy. This study presents the details of how the phases of the ADAPT-ITT framework were applied to the design of the adaptation. An advisory board was constituted from the target population, consisting of 6 African American women aged 18-24 years, participating in formative activities for 12 weeks, and involving components of the PMI design. As Afiya included a telephone-based PMI, developers of the original Afiya phone scripts crafted the initial design of the SMS-based texts and texting protocol. The advisory board participated in the 1-day Afiya workshop, followed by 4 weeks of texting PMI messages and a midcourse focus group, followed by 4 more weeks of texting PMI messages, ultimately ending with a final focus group. At the advisory board's request, this phase included an optional, additional week of text-based PMI messages.

Results: The methods provided a rich source of data and insights into the fundamental issues involved when constructing SMS-based PMI for this target population and for this EBI. Prior contact and context are essential as the health educator was identified as a key persona in the process and the messages were situated in the original (workshop) context. Narrative adaptations for personas emerged from advisory board discussions. Suggestions on how to expand the PMI to current, specific social contexts indicated that the use of narrative analysis is warranted.

Conclusions: The use of existing EBIs incorporating telephone-based PMI scripts facilitated the initial design of the texts, with a subsequent narrative analysis of the advisory board data providing additional adjustments given the actual context. Additional examination of the advisory board feedback revealed that personas would offer insight into and opportunities for a persona-specific modification of texting narratives.

(JMIR Form Res 2020;4(10):e22485) doi:10.2196/22485

KEYWORDS

short message service; HIV; African Americans; adolescent; female; texting; mHealth; ADAPT-ITT framework; intervention study; health status disparities; young adult; risk reduction behavior

Introduction

Young people in the United States have never known a world without HIV [1]. The National HIV and AIDS Strategy has consistently identified African Americans as a population at increased risk for HIV to which prevention efforts and resources should be directed (President Donald Trump closed the Office of National AIDS Policy in 2017) [2]. In recent Centers for Disease Control and Prevention (CDC) reports [3,4], in 2018, African Americans comprised approximately 12.3% of the US population but accounted for approximately 42% of the new HIV diagnoses and a disproportionate rate of HIV deaths at 28.4 per 100,000 population (vs 4.9 per 100,000 for Whites), leading all other racial and ethnic groups. African American females comprised approximately 13% of females in the United States but accounted for 58% of new HIV diagnoses among females, with 92% of infections attributed to heterosexual contact. The infection rate for African American females (21.3 per 100,000 population) was 13 times the rate for White females (1.7 per 100,000). These data underscore the HIV disparities disproportionately burdening African American women.

Disparities persist among adolescents as well. In 2018, the HIV diagnosis rate among adolescents in the United States was highest among African Americans (26.0 per 100,000), which was 19 times the rate for White adolescents (1.4 per 100,000) [5]. Young people aged 13 to 24 years accounted for 21% of new HIV infections, of which 52% were African American. The incidence among persons aged 13 years between 2014 and 2018 revealed that the rate among African Americans was still excessive, being 8 times that of Whites [6]. These disparities highlight a need for increased, effective efforts to address HIV prevention efforts among young African American women.

Challenges to HIV prevention include low testing rates, youth unsure of their HIV status, inadequate or delayed sex education, incorrect or no condom use, substance use, multiple partners, and high sexually transmitted disease rates. The 2017 Youth Risk Behavioral Surveillance Survey results from 144 high schools [7] estimated that African American adolescents, compared with White adolescents, were more likely (15.2% to 7.9%) to have been tested for HIV, more likely to have multiple partners (11.7% to 8.6%), and more likely to have had sex before the age of 13 years (7.5% to 2.1%). The required health education varied widely, especially within large, urban school districts where, across grades 6 through 12, the range was 0% to 100% for each grade (median: 6th grade, 59.9%; 7th grade, 47.6%; 8th grade, 27.0%; 9th grade, 75.9%; 10th grade, 44.4%; 11th grade, 39.1%; and 12th grade, 53.6%) [8]. Even when delivered in those districts, variances were found on coverage of all 20 topics in grades 6 through 8 (range 17.6% to 41.0%) and grades 9 through 12 (range 42.8% to 75.0%) [8].

A systematic review and meta-analysis of extant US school-based HIV and STI prevention programs found *no persuasive evidence for the effectiveness of school-based programs* [9]. However, a recent systematic review specifically examining the impacts on African American adolescents found that *sexual health interventions are associated with increases in abstinence and condom use and improved sexual health knowledge, self-efficacy, and intentions* [10]. It is essential that the response to HIV and AIDS effectively addresses African American youth.

Early prevention programs addressing behavioral risk factors are important for reducing the HIV disparity observed among African American adolescents relative to White adolescents. Although less is known about the burden and epidemiological details of sexually transmitted infections (STIs), especially HIV and AIDS, among youth and adolescents than about infants and adults [11], there is accumulating evidence that designing interventions for adolescents requires unique combinations and insight into gender, developmental level, race, culture, and social factors [12-14]. This is especially salient in the escalated concentration of HIV and AIDS in southern United States [3,15-17]. Understanding these factors in the context of adolescents is essential for designing and delivering effective prevention programs to ensure that they are developmentally appropriate [13,18]. In addition, there is one particularly important question that is salient to this paper: how can the effect of a successful intervention be maintained?

Program Maintenance Interventions

A review of national STI and HIV prevention programs targeting African American adolescents indicated that the *frequency and duration* of a program's implementation was related to its efficacy [19]. The implementation fidelity of extended face-to-face programs is difficult to maintain and expensive, particularly in an era of uncertain funding and support for HIV and STI prevention programs [20-23]. A recent metareview of youth-focused interventions concluded that they effectively influenced condom use, sexual health knowledge, and safer sex norms but suggested that future work *should focus on intervention adaptations and supplements that may extend their protective effects over time* [24]. Thus, the problem becomes one of sustaining program effectiveness over time—program maintenance interventions. This echoes the call from a focused review of HIV prevention interventions targeting African American women where technology can serve as *boosters* sustaining the effects of the intervention [25].

In general, preventive maintenance intervention (PMI) strategies such as making phone calls to participants have been more effective in sustaining intervention effects over time compared with interventions without a maintenance component [26]. However, incorporating intervention components to enhance behavior maintenance such as face-to-face programs and phone calls, is difficult to sustain, labor intensive, and costly [23]. Consequently, sexual health researchers have called for the increased use of technology-based interventions to promote the maintenance of HIV and STI preventive behaviors [27,28]. One

of the most promising technological platforms is the mobile (smart) phone delivering health-related services—texting.

The Social Role of Texting

Mobile health (mHealth) technology is gaining recognition in the field because it affords a wide array of user-facing functionality and services [29]. Evidence supporting the feasibility, acceptability, and efficacy of digital interventions is growing [30-32], particularly for adolescents [33-39]. SMS texting is seen as an mHealth technology with specific promise for adolescents [40], thus serving as an opportunistic technology for impacting sexual health behavior [41-45]. For younger adolescents (aged 13-17 years), texting is the preferred mode of communication with friends, surpassing social media apps, face-to-face interactions, and telephone calls [46], with 95% of adolescents (African Americans: 94%) having access to a smartphone and 89% of adolescents being on the web *almost constantly* or *several times a day* [47].

The critical point is that texting is a social medium. As opposed to theoretical stances that presume the *socialness* of technology is heavily determined by the characteristics of the channel's ability to provide cues, we adopted a social information processing perspective. This suggests that individuals adapt available cues for a given channel (eg, language, textual display options) to successfully accommodate their management of social interactions [48,49] over time. Technological characteristics are less about limiting the *amount* of social information exchanged and more about determining the *rate* of that exchange [49]. Furthermore, the social information processing approach is consistent with co-construction theories of adolescent development and technology-mediated exchanges, wherein adolescents are *cocreating the internet environment through processes of social interaction* and *construct the same developmental issues online as they do off* [50]. It is about establishing a common conceptual framework shared between the health educator and the client.

Adolescents' online and offline worlds are strongly connected as *they use online communication for offline issues, and to connect with people in their offline lives* [51]. Indeed, research on Twitter and Facebook has shown remarkable structural similarities between online social networks and offline (face-to-face) networks [52]. This could include direct mapping between the online and offline networks as well as the inclusion of online-only individuals, thus treating the modes of communication as *essentially the same* [52]. This explains the explosion of messaging among adolescents for relationship management. For adolescents, texting is as much a social choice as a technical choice. In fact, it is often the preferred choice [53], surpassing social media apps and phone calls, with adolescent females more likely to use texting as conduits for conversations with friends [54].

Consequently, we suggest that interventions beginning by establishing a strong social context (eg, initial meetings, workshops) would benefit from a texting-based PMI that can retain the social presence of that context by affording social presence cues to enhance the message [55]. The belief is that such hybrid intervention models can help sustain the impacts of EBI by sustaining the narrative engagement of participants. A mechanism we are developing to sustain narrative engagement is the tailoring of PMI messages via the specification of context-determined personas.

As with other intervention components, PMIs must be carefully designed and implemented, otherwise they may undermine the initial intervention [23]. Reviews of texting or mobile apps intended to increase various specific preventive or adherence behaviors in adolescents (eg, sexual health, smoking, oral health, sickle cell disease) reported that the overall efficacy findings were modest, but again only a minority reported a theoretical framework in their design [34]. The feasibility and accessibility of SMS as a behavioral maintenance strategy for African American adolescent populations with increased risk of HIV and STI acquisition requires *continued adaptation of evidence-based interventions with text messaging–enhanced content to expand our knowledge of the potential of this approach* [56].

Therefore, instead of starting with the technology, start with a theoretically grounded EBI, then discern how to *adapt* the EBI to technology to each other. The integrity of the EBI must be maintained whereas the capability of the technology must be exploited in the context of the target population. Although a variety of *generic* approaches (design guides or review recommendations) for integrating technology into behavioral change interventions have been proposed [57-62], a more focused approach was appropriate for this adaptation. We selected a design framework to guide how the *intervention can besystematicallyadapted to a technological form and use*.

In this paper, we present a formative pilot study that describes the design process applied to adapting an EBI to reduce the risk of HIV and STI among African American adolescent females. We selected an existing EBI for this target group that would likely benefit from a text-based program maintenance intervention to extend the EBI's reach, sustainability, and effectiveness. We applied an adaptation-implementation framework that demonstrated efficacy for that purpose—the Assessment, Decision, Adaptation, Production, Topical experts—Integration, Training, Testing (ADAPT-ITT) framework [63].

ADAPT-ITT Framework and Implementation Science

Calls to advance implementation science argue for the development of adaptation strategies *that would more comprehensively describe the needed fit betweeninterventions and their settings* [64]. A recent review [65] traces the history of the most prominent models, including the ADAPT-ITT framework, noting the primary health context of their application and extracts steps that is common across frameworks. These frameworks reflect the CDC's earlier general guidance describing the ADAPT process intended to assist health departments and community-based organizations in adapting an EBI *to fit the cultural context, risk determinants, risk behaviors, and unique circumstances of the agency without competing with or contradicting the core elements and internal logic* [66].

The purpose of this research is to use the ADAPT-ITT framework to modify existing EBIs to increase STI and HIV

preventive behaviors among African American females. Adaptation focuses on translating health educator–facilitated phone calls to SMS-delivered messages. We selected the ADAPT-ITT framework because of our experience with the framework, its relevance to our target population, and its demonstrated effectiveness in guiding the design implementation of postworkshop PMIs to extend elements of the primary intervention over time [67]. The ADAPT-ITT framework is being widely used to design and implement HIV and STI prevention interventions and other prevention adaptations in diverse settings in different cultural contexts, both domestically and internationally [68-78].

Methods

This was a formative study using qualitative methods. ADAPT-ITT has 8 phases. Each phase brings contextual nuances and constraints that determine how the phases are engaged and the timing of corresponding tasks. The phases are as follows: (1) Assess the proposed new priority population's HIV risk profile, (2) Decide on whether to adopt or adapt an EBI, (3) Administer novel methods to facilitate the adaptation process, (4) Plan on what aspects of the EBI need to be adapted and plan on how best to evaluate the adapted EBI, (5) identify Topic experts to assist in the adaptation process, (6) Integrate material from the topic experts to adapt the EBI, (7) Train staff to implement the adapted EBI, and (8) Test the adapted EBI. Our texting adaptation process describes how the ADAPT-ITT phases can be used to guide the adaptation of EBI components to texting.

The goal was to select an existing EBI for African American adolescent females who would benefit from text-based PMIs to extend the EBI's reach, sustainability, and effectiveness. We applied the ADAPT-ITT framework to translate an original EBI (Afiya) that included a face-to-face workshop followed by health educator phone call PMIs to an adapted intervention that would include the same face-to-face workshop but followed-up by health educator texting PMIs. Table 1 summarizes the 8 steps and the lessons learned in the context of this adaptation pilot. The lessons learned identified as *structure* describe changes made to EBI delivery or logistics. The lessons learned identified as *content* describe changes to the information delivered in the intervention. The core element of the translation process is the formation of an advisory board (advisory board) selected from the target population.

Table 1. Application of the Assessment, Decision, Adaptation, Production, Topical experts—Integration, Training, Testing framework with a preventive maintenance intervention texting adaptation in Afiya.

Phase	Methodology decisions	Results or observations
1. Assessment	• Formed advisory board • Conducted an initial meeting with advisory board • Conducted meetings with intervention experts for the target population	• The target population (African American adolescent females) requires interventions addressing their specific situation • The intervention needs to address safer sex norms, sexual negotiation, and refusal skills, HIV and STI[a] preventive attitudes • The advisory board was selected for project, met, and provided input on the plan • Advisory board members were paid as consultants to develop or test the feasibility of texting PMI[b]
2. Decision	• Selected HORIZONS and Afiya interventions • Decided to adapt Afiya phone-based PMIs to texting PMIs • Original developer of Afiya-led adaptation • mHealth[c] and texting literature reviewed	• HORIZONS, an evidence-based intervention that successfully reduced STIs and increased condom use among young women, was selected as it appropriately addressed the target population • Afiya selected as it demonstrated a PMI based on HORIZONS to the same target population, but extend the efficacy of the HORIZONS intervention • Texting is most likely the simplest mHealth adaptation to adapt from Afiya and appropriate implementation technology choice for the target population
3. Administration	• Decided not to modify Afiya core elements • Decided to focus on Afiya core element support by texts • Adaptation would involve scripts and texts responses based on Afiya documents and data	• Retention of primary intervention form assured standard of care and served as a shared primary intervention control (therefore, theater testing was not conducted) • 5 core elements of Afiya were identified to be supported by PMIs
4. Production	• Designed 8-week iterative formative pilot plan for adaptation • Designed questionnaires and focus group–led questions • Designed initial texts and support scripts • Selected company for formative pilot texting • Reviewed with Afiya and HORIZONS creators • Reviewed plan with advisory board • Assembled Afiya workshop materials	• Iterative formative pilot designed for advisory board • Primary intervention led by Afiya and HORIZONS developers in formative pilot and support scripts and associated text messages were created • Important to have both advisory board and intervention developers review pilot design and timing • Important to have intervention developers review PMI scripts and texting form or content (Draft 1 completed) • Texting platforms vary in terms of capacity, functionality, and cost. Total program costs should be estimated
5. Topical experts	• Obtained review advice regarding HIPAA[d] issues relevant to texting • Obtained review advice regarding pre- and postintervention CASI[e] instruments for future effectiveness pilot • Met with texting companies (with advice from IT[f]) for future effectiveness pilot	• CASI content should be reviewed for duplication or replication as well as total time required (test run the instrument) • Texting is not encrypted, so both HIPAA (and local IRB[g]/state) requirements must be reviewed regarding texted data and data privacy issues • Role, functionality, and cost of texting companies were more clearly defined with this experience • Overall costs for the 8-week formative pilot were made
6. Integration	• Integrated Draft 1 and topical expert comments for scripts, SMS texting content or form (Draft 2) • Conducted Afiya workshop with advisory board • Integrated topical expert comments of pre- and postworkshop CASI materials • Integrated Draft 2 and CASI components (Draft 3) • Final review by advisory board and Afiya and HORIZONS creators	• Draft 2 completed • Draft 3 completed • Final formative pilot design completed
7. Training	• Created training materials for the formative pilot	• Formative pilot materials completed
8. Testing	• Submitted to IRB for clearance of the formative pilot	• Obtained IRB clearance to determine adaptation efficacy

Step 1: formative pilot

Phase	Methodology decisions	Results or observations
	• Conducted 8-week formative pilot with advisory board: primary Afiya intervention, 4-week texting, focus group 1 (adjustments), 4-week texting, focus group 2 (adjustments)	• Iterative pilot design provided important feedback on texts (timing, wording, dose) as well as script and CASI modifications • Information gathered during the first advisory board focus group was applied to modifying the second 4-week texting session, information gathered during the second 4-week focus group was also useful (topic experts were consulted during these sessions as necessary) • 5 personas identified, exit interviews added • Results informed design of adaptation impact pilot
Step 2: feasibility pilot		
	• Designed texting PMI RCT[h] feasibility pilot study • Study methods submitted to IRB for approval	• Obtained IRB clearance for 3-arm feasibility pilot study • Preparing for PMI RCT feasibility pilot study to examine short-term intervention efficacy

[a]STI: sexually transmitted infection.

[b]PMI: preventive maintenance intervention.

[c]mHealth: mobile health.

[d]HIPAA: Health Insurance Portability and Accountability Act.

[e]CASI: computer-assisted self-interview.

[f]IT: information technology.

[g]IRB: Institutional Review Board.

[h]RCT: randomized controlled trial.

Phase 1: Assessment

Who will be the primary audience for the EBI? Phase 1 included incorporating information from the published literature on the need for behavioral maintenance strategies that are sustainable, encouraging health behavior maintenance (as previously summarized). The epidemiological data for this target population highlighted the health disparities in the prevalence of HIV transmission through heterosexual contact among African American female adolescents and young women. Our target population was well defined, and the population attributes and risk profiles were already matched with EBI theory and design approaches to interventions. Therefore, we could immediately recruit an advisory board from that target population. The advisory board members consisted of 6 African American women aged 18-24 years to serve as additional assessment-informative roles throughout the adaptation process. The advisory board members were compensated to participate in the formative advisory activities for 12 weeks.

Phase 2: Decision

What EBI Will Be Used and Will It Be Adapted?

In phase 2, the critical decision on which EBI to select or to determine whether any EBI would be appropriate is made. This necessitates an examination of EBIs and of how likely the target population is to respond to the adapted components of the intervention. In this case, texting. This provides the theoretical and empirical foundation for the form and substance of the texting PMI.

Selection of the Primary EBI: Afiya

We examined several HIV and STI interventions explicitly designed and tested for either Blacks, Black females, Black adolescents, or Black adolescent females. Afiya was selected as the intervention to be adapted for the project for 3 reasons. First, Afiya [79] incorporates the HORIZONS intervention that has been evaluated and designated a tier I (best) evidence-based risk reduction intervention by the CDC [80]. HORIZONS, an in-person group-delivered session tailored for Black adolescent females, significantly reduced chlamydial infections and increased HIV and STI preventive behaviors, partner communication efficacy, condom use efficacy, and HIV and STI prevention knowledge. Importantly, the underlying theories of HORIZONS and Afiya are social cognitive theory [81] and the theory of gender and power (TGP) [82-84]. The TGP addresses partner influences and gender-based social correlates that influence behavioral risk factors, which have been extended to address the exposures, social or behavioral risk factors, and biological properties that increase women's vulnerability to HIV acquisition [83,85].

Second, Afiya used HORIZONS plus bimonthly phone calls for 3 years to determine whether healthy behaviors could be maintained over time. The calls were brief, tailored one-on-one sessions with a health educator, delivered over a 36-month period, and the calls served as PMIs. In the Afiya trial [79], the workshop (tested in HORIZONS) was implemented as a single 4-hour group session delivered by trained African American female health educators. When answering a call, health educators followed an algorithm in their call scripts to guide them through scenarios involving intimate and sexual partnerships, referencing information from the workshop. Afiya (1 workshop+18 phone calls) was efficacious. Over the 36-month follow-up, compared with the control group, participants in the treatment group had significantly lower incidences of chlamydial (50% reduction) and gonococcal infections (60% reduction), a higher proportion of condom use (both within 90 days and 6 months before assessment), and fewer sexual episodes when high on drugs and/or alcohol [79].

These findings provide initial insights into (1) how to begin to adapt the telephone-based PMI to an SMS-based PMI and link them to core elements of the intervention theory and (2) how the target population would respond to a communication mode that differed from the initial telephone calls.

Third, the developers of both HORIZONS and Afiya were available for consultation and reviewed the project products to maintain the integrity of the adaptation and ensure that it remained faithful to the theoretical components of the original intervention.

Selection of Texting Adaptation

As discussed, simple mobile phone–based texting was selected as it is ubiquitous, tethered to individuals, and a widely accepted technological application, especially in the target population (teen or adolescent African American women). Using mobile phone technology for adolescent and young adult health purposes was promising as most members of this population possess this technology and exhibit distinct patterns of use, especially the enduring dominance of texting [40,86,87]. Thus, in terms of reach, texting affords a *preimplemented, preadopted, and preferred* technological platform for adolescents.

Phase 3: Administration

What components of the selected EBIs should be adapted? We decided that the adaptation should address the selected core elements of the EBI: (1) ethnic and gender pride role models, (2) sexual health decision making, (3) HIV and STI knowledge, (4) healthy and unhealthy relationships, and (5) negotiating safe sex. It is important to note that several of these activities involve direct, face-to-face, and small group interactions (discussion and simulations) to address situational-partner negotiation strategies and skills, and other activities involve physical skill development and practice (eg, condom application). As these are both essential and not substitutable by texting, it was determined that the adaptation would not be directed toward modification or replacement of the activities associated with the core elements of the workshop. Discussions with experienced health educators noted that each core element and its associated activities supported thematic narratives that recurred in the workshop. Consequently, the core elements of the EBI did not need to be adapted; rather, the narratives that surrounded each core element needed to be extended. Texting, as it turns out, is a powerful, yet efficient, mechanism to accomplish this. Toward that end, all messages would be set in the context of participants in interacting with messages prepared by Tina (one of the health educators).

The adaptation of Afiya involved modifying the telephone scripts and phone-based messages to an SMS text narrative form, aligning with the Afiya core elements. These messages were designed by the developers of Afiya and served as plausible *first approximations* to test the messages, the messaging concepts, and the messaging technology with members of the target population in the formative pilot (phase 8). Thus, the Afiya scripts were predefined narrative structures tailored to the texting form (examples shown in Table 2). As this is a text-based intervention, theater testing was not used; rather, equivalent methods (and supporting theory) designed to examine how humans interact with computers were employed. In general, this is referred to as user experience (UX) analysis and design [88,89].

Table 2. Mapping Afiya core elements to texting categories.

Afiya core elements	Primary intervention activities	Associated text category or example
• Ethnic and gender pride • Role models	• Poems and affirmations • Media images • Role models	*Poems and affirmations (motivational message):* "It is where you are headed not where you are from that will determine where you end up (Marion Wright Edelman)"
• Sexual health • Decision making	• Goal setting • Values • Sexual health options (condoms, abstinence)	*Sexual health choices (vote poll):* Question: What is ur sexual health choice? • Abstinence • Condoms • Neither • Not sexually active Automated responses: • Abstinence is the 100% way 2 protect against STDs[a]/HIV. Remember 2 talk with ur partner abt ur sexual health choice • Remember 2 use condoms each & every time u have sex. Talk with ur partner abt ur sexual health choice before u have sex • Abstinence is 100% STD[a] protection. Condoms protect u from STD if u use them each&every time. Consider 1 of these options • If u decide 2 have sex, condoms protect from STDs if used each & every time. Talk with ur partner abt ur choice
• HIV/STD knowledge	• Facts about STDs • Testing • OPRaH • Condom do's and don'ts	*STD testing (cues to action):* Question: When do u need an STD test? • Partner has STD • Had sex-no condom • New partner • All of the above Automated responses for (A)-(D): Get tested when new partner, STD symptoms, sex no condom, ur partner has STD. Planned parenthood MTW&F[b] 8-4:30pm; xxx-xxx-xxxx
• Healthy and unhealthy relationships	• Personal risk factors • Understanding risks • What turns you on? • Boundary setting • Peacefully breaking up	*Boundary setting (cues to action):* Question: Y is it important 2 know ur sexual health boundaries? • 2 help stick 2 ur limits • 2 communicate them to ur BF • A & B Automated responses: • One that's a start! Stick 2 ur limits. Its okay 2 say what ur comfortable doing & what u don't want 2 do • That's a start! Telling ur partner what ur boundaries r is important so he doesn't make u uncomfortable or cross the line • Exactly! U want 2 b comfortable & happy in ur relationship. Communicating ur boundaries 2 ur bf is healthy
• Negotiating safe sex	• Healthy communication • Sexnarios: negotiation skills • Condom excuses and comebacks	*Communication (cues to action):* Question: What type of communicator are u in ur relationship? • Aggressive • Passive • Assertive Automated responses: • If u r aggressive, do 1-2 assertive things. Avoid yelling & threatening. B direct & honest. C how he responds. • Passive If ur passive, do 1-2 assertive things next time. Avoid the silent treatment. B direct & honest. C how he responds. • Great! Continue 2 b assertive when u talk 2 ur partner, even when it gets hard. Be direct & honest. C how he responds

[a]STD: sexually transmitted disease.

[b]MTW&F: Monday, Tuesday, Wednesday, and Friday.

Text messages fit into 1 of 3 categories: motivational messages, vote polls, and cues to action. Motivational messages are one-way messages delivered to participants to serve as inspiration using poetry from Black authors and other leaders.

Vote polls are two-way messages initiated by Afiya's texting platform. The vote polls pose a close-ended question, and participants respond with the letter corresponding to their answer. On the basis of their answers, the system has predesigned responses that will automatically reply to the participant. The automated replies provide information linked back to the Afiya content, serving as a reminder of key messages from Afiya. The cues to action are like vote polls and are two-way messages initiated by the Afiya texting platform. However, cues to action are specifically designed to have automated replies that cue or prompt the participant to engage in a specific action, such as sexually transmitted disease testing or initiating a conversation regarding STI prevention with a sex partner. Each message maps on to at least one of the Afiya core elements presented in the main intervention workshop.

Phase 4: Production

How do you adapt and document adaptations? We analyzed the results of the focus groups and texting and then reviewed them with the HORIZONS and Afiya creators. Initially, the responses to SMS texts were made by a health educator and guided using the modified Afiya phone scripts, and these were adjusted based on the analysis and the consultation with the HORIZONS and Afiya developers. We also initiated the design of computer-assisted self-interview (CASI) questions and instruments as well as focus group–led questions. Measures of fidelity and quality assurance included postsession interviews and participant evaluations. Initial designs of the SMS texting and automated two-way communication were created with help from the Afiya developer with respect to the core elements of the EBI.

In parallel, we began to design the implementation requirements for selecting a commercial texting platform that would best serve our formative texting pilot needs to capture data from a research study, including questionnaires, instruments, and focus group–led questions. At the time of the study, many commercial texting platforms were built for mass marketing. We needed a platform that would allow us to easily set up and automatically send messages based on their cohort or study condition as well as conditional texting responses given a texting question (probe). Additionally, we needed to be able to initially use the platform for recruitment and then transition people into the appropriate study cohort or condition once they consented and were enrolled into the study. Once selected, we would have to code the scripts or texts and conditions as required by the vendor to realize the texting PMI for Afiya.

During this phase, the group drafted an 8-week formative pilot plan to engage the advisory board in the adaptation process. However, given the nature of mHealth technology and uncertainty regarding the specifics of adapting the narratives, we engaged an alternative design method involving an individual texting and response analysis, followed by a group discussion. This resulted in iterations between the administration and production phases, reflecting a development approach often used in software projects to define and refine the UX by tightly integrating interactions and adjustments in design [90].

The overall sequence of the formative pilot is as follows: advisory board members initially participated in a 4-hour HORIZONS workshop (ie, the primary intervention) led by a health educator. Following the workshop, weekly automated SMS text messages were sent that highlighted key messages from the core elements of the workshop for 4 weeks. Advisory board members would activate the weekly text by responding to a weekly text prompt, such as:

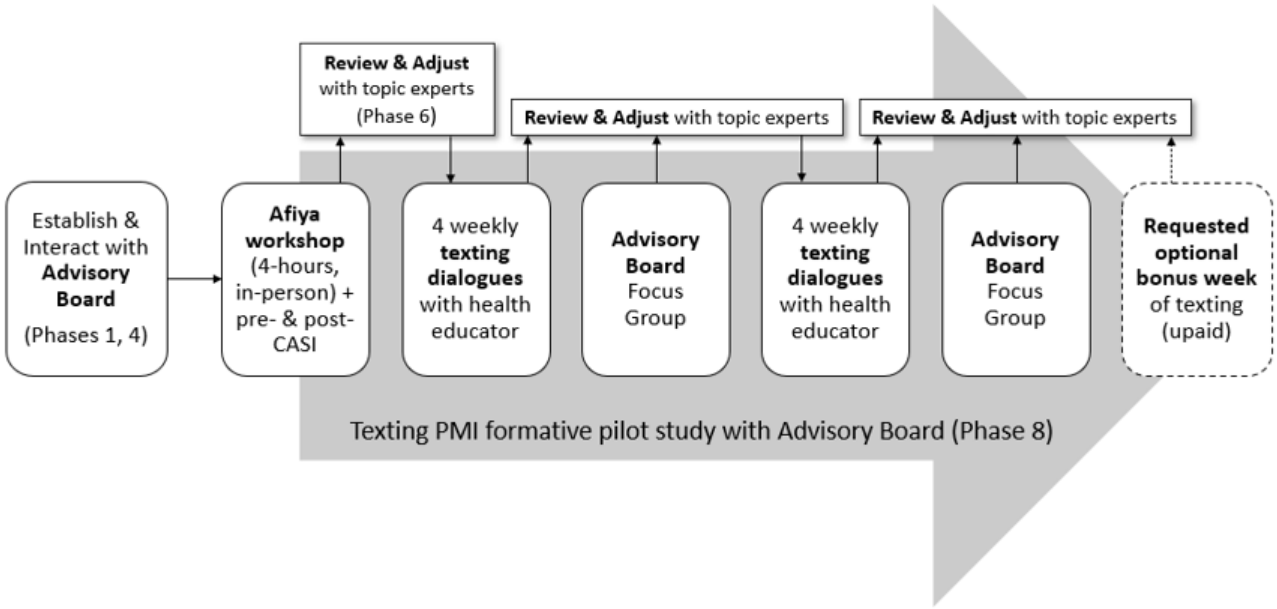
> *Good morning! 2 get ur next HORIZONS convo with Tina, reply to this text with the word WEEK1*

They would then receive a text that required a response related to a specific core element covered in the workshop, such as sexual health decision making:

> *Afiya talked abt sexual health choices: abstinence, condom use or neither. What choice is best for u right now?*

Subsequent texts were selected using response scripts based on the original phone scripts used in the Afiya phone-based PMI. The timeline for the pilot study is shown in Figure 1. At the end of 4 weeks, the advisory board focus group met to provide feedback. The team reviewed the results and adjusted the scripts and texts. There were then 4 additional weeks of messages followed by the final advisory board focus group, allowing 2 cycles of iterated user response elicitation and refinement (this type of small-scale, iterated focused development is common in UX studies and is similar in theory to spiral development in software engineering but involves rapid testing and turnaround. The presumption is that users may have opinions of their needs and how they would react to a system but after actually engaging a small prototype system the requirements change, which will cause the system to be modified and reassessed in an evolving spiral that converges as the system is developed. This is most common when requirements are ill-defined, as is often found when implementing new types of interactions with technology). Finally, the decisions, designs, and information emerging from phases 1 to 4 were integrated into the first draft, Draft 1.

Figure 1. Timeline for formative pilot study. PMI: preventive maintenance intervention.



Phase 5: Topic Experts

Who can assist with the adaptation? Although experts were consulted as needed throughout the process, phase 5 involved additional experts whose input was specifically necessary to help review the adapted intervention design (content and form). In this case, additional experts were solicited: (1) experts in designing CASI items and psychosocial instruments relevant to the purpose and population, (2) representatives from selected third-party texting companies to determine capabilities, constraints, and pricing, and (3) those who had specific Health Insurance Portability and Accountability Act knowledge regarding ethical-legal implications of the possible SMS texting content and data privacy or protection issues. We generated initial estimates of total program costs and had them reviewed by the team and administrators. Primary cost estimates included the texting company, a graduate research assistant (GRA), and participant incentives.

Phase 6: Integration

What is going to be included in the adapted EBI? Phase 6 involved the integration of the results from Draft 1 (phase 4) with the results of the topic experts' review of the fidelity of the script or message designs relative to the original Afiya intervention core components (phase 5) to create Draft 2. Project members reviewed the pre- and post-Afiya workshop CASI measures (the workshop was offered when phase 4 CASI materials were completed and vetted) and revised the instruments to (1) reduce their length and (2) include assessment items appropriate for evaluating an SMS-based intervention. We finalized the modified Afiya scripts for a texting algorithm that would be used by the health educator in the postworkshop texting in phase 8. These were reviewed by Afiya developers and the advisory board. Our team produced Draft 3 after completing the workshop and after text communication with advisory board members in phase 6.

Phase 7: Training

Who needs to be trained? Once the decisions were made regarding the adapted text intervention (scripts, SMS texts, texting platform), intervention-specific training materials were developed for the formative pilot. Health educators were trained on how to implement the Afiya workshop; materials were already available, with minor modifications introducing SMS texting and approved by the Afiya developers. Health educators were trained on the SMS text scripting component and texting platform use and were part of the development process. One of the study principal investigators and a GRA were trained to develop and manage the texting platform and messages. Data analysts were trained on the texting platform's data components such that texts and associated data could be acquired from the company.

Phase 8: Testing

Was the adaptation successful? The definition of success and the form of the testing depend on the context of the adaptation, but the general objectives remain. As our adaptation did not include delivery to a different target population, we modified the two-step testing processes of phase 8.

In Step 1, we assessed the *adaptation efficacy* by conducting an 8-week formative pilot study with the advisory board members, as shown in Figure 1. This important validation step addresses whether the intervention was effectively adapted and implemented correctly, involving close monitoring and specific types of data acquisition and feedback and testing all the operational components including health educator training and performance of the texting company. We decided against a *pure* formative pilot (ie, engaging a new set of participants) because of the extensive involvement of the advisory board and HORIZONS and Afiya developers, the retention of the same target population as Afiya, and the success of the iterative design process with the advisory board for the formative pilot. This approach was extremely helpful and informative in designing

the final product of this stage, the randomized controlled trial (RCT) efficacy study.

Step 2 involves assessing the *short-term intervention efficacy* of the PMI by conducting an *RCT feasibility study* to test the efficacy of the PMI over an extended duration with the target population. Regarding Part 2, a 3-arm RCT efficacy study (standard of care, Afiya workshop only, and Afiya+texting) was designed. This has been approved by Emory's Institutional Review Board and is being initiated.

Results

As noted, the formative study involved the advisory board members (n=6) actively participating in adapting Afiya to a texting platform, and they were compensated for their time and effort. Focus groups with the advisory board were conducted midway through the pretest and after the last week, allowing 2 cycles of iterated user response elicitation and refinement. Examples of the feedback after week 4 are summarized in Textbox 1. For this focus group, we sought to distinctly examine 4 categories of texting issues: timing, wording, content, and form. The narrative analysis was very useful in extracting issues regarding texting characteristics, such as timing (eg, *thank you* messages came too early in the texting series), wording (eg, questions were judged as being clear and simple), content (eg, inspirational messages were appreciated), and form (eg, desired more individualized feedback). Getting the text wording *correct* for the target audience was also a continuous improvement process to facilitate culturally appropriate narrative engagement via text. Frequent use of commonly agreed-upon abbreviation dialects in texting results in these abbreviations being rapidly absorbed into the language and automatically (linguistically) recognized as *words* [91], reflecting how humans socially adapt to a technology (and its constraints). On the basis of the first focus group with the advisory board, the following modifications were made for the second 4-week session:

1. We added a motivational or inspirational message each week in addition to the regular text, such as "It is where you are headed not where you are from that will determine where you end up" (Marion Wright Edelman).
2. We added more complex response scripts to the prompts because participants felt some of the texts were too open ended and not specific to them.
3. We provided individualized feedback and advice during a time for open questions about any topic.
4. We gave advisory board members a specific window of time to chat with the health educator
live
instead of always getting the automated message responses.

Textbox 1. Results from the advisory board texting formative pilot (texting issues and examples of reactions; weeks 1-4).

Timing:

- Texting frequency should be increased; daily texts are acceptable
- No preferences for weekday vs weekend texts
- Texts should not come too early in the day; better to get texts starting mid-day and continue through early evening
- Participants indicated that the Thank You message came too early in the text series and would have preferred a more in-depth conversation via text

Wording:

- Text wording was straight-forward and participants liked that the messages were not *sugar coated*
- Questions were clear and simple

Content:

- Texts about sexually transmitted diseases were favorite and made them think about their behaviors
- When discussing relationships, participants emphasized that content should focus on communication in relationships and engaging in safe behaviors. They also wanted the discussion to expand to other facets of relationships (eg, quality of relationship, satisfaction with relationship)
- For more complex issues like relationships and communication, the participants requested more feedback and information to help keep them on track
- Participants indicated that any topic would be acceptable by text if they knew where their answers were going and who was reading or responding to the texts
- Wanted a motivational or inspirational message or quote each week

Form:

- Dissatisfied that there was not a person live on the other end of the texting
- Disliked texting with a computer that could not be more interactive and have a conversation with a live person
- Wanted more individualized feedback and advice to come with the texting program

All advisory board members participated during weeks 5 and 6; participation decreased during the second half of the pilot. Reasons for nonparticipation were as follows: (1) phone temporarily disconnected and (2) busy during the specific designated conversation window (the scheduled time to interact). Within the final focus group, comments were sought regarding their overall experience and thoughts for the PMI. Examples of the feedback after week 8 are summarized in Textbox 2. These fell into 4 primary categories: HIV and STI information, context-specific social network maintenance, social engagement, and other ecologically relevant information or content. Participants sought additional information on related programs and topics beyond those addressed in the workshop. They also reported that the social connection with the health educator was indeed maintained, facilitating engagement. An interesting consequence of the pilot was the value of the social network that emerged within the context of the pilot and the role that delivery platform could play in maintaining that social network.

Textbox 2. Results from the advisory board formative pilot (weeks 5-8, final focus group).

HIV and sexually transmitted infection information:

- Participants wanted to make sure that information about new programs they may be able to participate in would be sent to them
- Participants requested that the texting program also be a forum to let participants know about group meetings and group events surrounding related topics, such as medical or health service options (clinics that provide free or reduced-price services), education events, and training activities

Context-specific social network maintenance:

- They stated that they enjoyed the texting but loved the groups
- They emphasized repeatedly a desire to continue meeting (in person) with the groups
- They expressed a design to have a forum to stay in touch in case their phone numbers change
- Participants noted that they liked having greater time to talk and the ability to have more detailed, back-and-forth conversations

Other ecologically relevant information:

- Participants stated they wanted information on the following topics: relevant community events, parenting, housing (different locations, pricing, etc), job information (job fairs, career info, etc), and health insurance information

Social engagement:

- Participants stated that they felt like they *were texting with a friend* and felt very comfortable texting with the health educator. Important characteristics for health educator: woman, relatable, trustworthy, the same person over time to develop a relationship with, knowledgeable person who can provide accurate information

Postadaptation, we analyzed how advisory board members engaged during the adaptation process, focusing on the emergent themes of interactions, to define an initial set of personas. Personas are specific, but hypothetical, individuals possessing core sets of relevant characteristics (eg, situations, beliefs, goals) that differentially influence the interaction narratives. The persona analysis defined 5 personas and linked each persona to counseling scripts from the Afiya intervention that reflected the context of prevention needs based on the situational narrative, the condom use behavior, and the current status of any relationships. Persona results were reviewed and revised by Afiya health educators and the project team. Note that personas are usually stable (as a descriptive model for the target population), but individuals may move between personas. If certain characteristics change, persona classifications can be modified (eg, a *Shannon* may turn into a *Lexie*). The results are shown in Table 3.

Table 3. Results of the persona analysis for the Afiya texting adaptation.

Characteristics	Shannon	Tonya	Kia	Shauna	Lexie
Narrative	Shannon has been in a long-term relationship with her boyfriend for almost a year. They stopped using condoms after the first couple of months	Tonya is single. She just got out of a relationship and needs some time to focus on herself and her schoolwork	Kia is dating someone, but also *talking* to a couple of other guys. She does not have a formal boyfriend but has a consistent sex partner. However, they are not in an exclusive, monogamous relationship	Shauna is starting a new relationship with a guy. They've been seeing each other for a month. She is abstinent and is ready to let him know her sexual health choice	Lexie has both a boyfriend and a casual sex partner
Condom use	Shannon has discontinued the use of condoms	Tonya is not having sex right now but wants to use condoms when she does get into a relationship again	Kia uses condoms with her sex partners	Shauna does not need condoms at this time as she is not having sex	Lexie uses condoms with her casual partner but not with her boyfriend
Relationship status	Shannon has had a boyfriend for 1 year	Tonya has no boyfriend	Kia has no boyfriend but has multiple partners	Shauna has a new partner	Lexie has concurrent sex partners
Link to Afiya intervention scripts	Script D: no to condom use	Script A: no boyfriend; me time	Script C: yes to condom use	Script B: yes to abstaining	Script E: inconsistent condom use

Finally, upon request from the advisory board members, we provided them with the opportunity to text in with any questions they may have. This was an opportunity for them to seek additional information that they may not have learned through the program thus far. The text lines were open for 7 days. advisory board members were not compensated for texting during the bonus week 9; it was an opportunity for them to get any additional information from a health educator before the program ended. Half of the advisory board members participated in the unpaid bonus week of texts. The chat line parameters were as follows:

1. Chat lines open for 7 days.
2. You will not get paid for these texts.
3. You can ask about anything you have questions or concerns about.
4. You will get a response from Tina within 24 hours.
5. If you have an emergency, you should always call 911.

Discussion

The motivation for this paper was provided by the significant disparities in STI and HIV health risks among African American youth. Ethnic minorities in the United States have STIs 30 times greater than that in White, middle-class populations [92]. To address this disparity, coupled with scarce funding and staffing resources, new forms of digital technology solutions are constantly being sought, proposed, and attempted to curb the STI and HIV risk behaviors of adolescents and young adults, who are at most risk [67]. However, few guidelines currently exist to inform designers how digital media can be used in ways that can impact the efficacy of an intervention.

This paper describes a conservative approach to contribute to this effort. Specifically, this paper describes the development of a systematic process to incorporate simple texting messages as an adjunct program maintenance intervention to an evidence-based sexual health intervention for African American adolescent women.

The systematic approach of the ADAPT-ITT framework was clear and effective in phases 1 through 7. The reason was the reliance on the well-proven and understood primary intervention (HORIZONS) and the prior efforts (and developers) experienced in adapting that intervention to a communication-based PMI (Afiya). However, in developing technological implementations, we found it necessary to engage a sample of the target population with the PMIs directly, rather than the theater-testing approach suggested in phase 1.

The type of small-scale, iterated development approach with the formative pilot is common in UX design and software engineering. The presumption is that users may have *beliefs* about their needs and how they would react to such a system, but after *engaging* a system, many of their beliefs and preferences (therefore the requirements) change. This is most common when implementing new (to users) types of interactions with technology. Consequently, the formative pilot was conducted in phase 8: Testing, as it required a full implementation of the PMI to gather efficacy and implementation data.

As shown in Figure 1, the formative pilot included 2 focus group sessions interspersed within the SMS texting intervention. This allowed intermediate analysis of the 4-week experience to be reviewed by the developers (and any relevant topic experts) and adjust the SMS texting narrative characteristics. Thus, the formative pilot testing in phase 8 *looped back* to elements of phases 4 to 7.

It is unlikely that technology will elevate an ineffective intervention to one that is effective—automating a weak process generally results in an automated weak process. There is also a risk that poorly designed and implemented technological components may lower the efficacy of a proven intervention. The selection of texting was seen as a key technology commonly adopted by this group as a mechanism for sociocultural connectivity, thus having the potential to increase the uptake of extant interventions by extending the narratives that began in the core EBI workshop activities.

Early and repeated inclusion of the target population, advisory board, was essential in the development. In particular, the generation of personas from the advisory board participants provided insight into important narrative variations for subsequent delivery design. Key considerations for end users should be sought early in the process of app (or other digital) behavioral intervention design to ensure both short- and long-term engagement [93-95]. In the paper by Pettifor et al [13], titled Adolescent Lives Matter, they argue that "adolescents should be involved throughout the process from design to implementation." This is supported by a recent systematic review of mHealth apps for adolescent users and concluded that there is a fundamental difference between adult and adolescent preferences in apps, where apps for adolescents need features that support decision making [91]. In summary, a quote by Glasgow et al [96] echoes the spirit of this paper's intent:

> *Controlling the epidemic will not only require bringing successful interventions to scale but also tailoring them to vulnerable and marginalized populations and understanding the social, cultural, and institutional contexts in which interventions are delivered.* [Page S26]

In summary, 4 general suggestions can be made regarding how to proceed when considering mHealth solutions.

Suggestion 1

Do not start with the technology. Start with the EBI-based workshop and then discern how to adapt both.

EBIs (and the underlying theoretical base) define core elements that have demonstrated effective results. Technology alone cannot *increase efficacy to significance* but has the potential to decrease efficacy to insignificance. This leads to the necessary inclusion of methods that address user-centered issues when designing the technological components.

Suggestion 2

Technological adaptations of EBIs need to explicitly address the user's experience as developers design, evaluate, and implement the technical components of an intervention.

Participants of technology-based intervention components are indeed *users* interacting with a service. Much is known about how to design, evaluate, and implement user services in general and how to assess the UX with technology in particular. This leads to an important conclusion regarding how to view the UX with adapting an EBI using a texting PMI.

Suggestion 3

Texting is less of a technical choice and more of a social choice.

Texting is not about creating a technological intervention and then trying to get individuals to adopt and to disseminate it. Everyone who has a mobile phone has the application; most people know how to use that application; most people have adopted that application for use. There is another element to consider; people use it for the same reason: *engaging in narratives with others*. Thus, a critical communication pathway already exists but it is limited. Interventions that use mHealth and texting compete for the user's attention and the user's time. Regarding the EBI, attention is gained, for example, via the EBI workshop (as demonstrated for this population for this health problem).

However, the narratives occurring in the workshop must be maintained through the initial (and continuing) establishment of *trust*, which mediates between information quality and use [97]. All health educators know this. By understanding the cultural components of the target population's preferences, choices of how the technology is used can engender that trust, thus sustaining critical narrative connections. Indeed, today's young adults value technology as *a way to enhance, not replace* their interactions with their health providers [37]. This leads to our fourth suggestion.

Suggestion 4

Texting as a PMI used with an EBI workshop is less about reminding and more about re-engaging narratives from the core intervention.

Texting can be used to extend the information and skills covered in an initial workshop or session. The Afiya intervention did not cover new information via texting. All information was presented in the in-person initial workshop. Therefore, testing then served to boost and further highlight details covered in the initial intervention. For example, the initial Afiya workshop demonstrated communication and condom use skills. The follow-up texts used terminology from the original intervention to remind or cue participants to the content. Subsequent text messages would then ask the participant about her communication style or whether she talked with her partner about her sexual health choice. The use of SMS in behavioral research should consider having an initial in-person presentation of information and using SMS to serve as a reminder of the information, resources, and skills presented.

Limitations

The challenge in future research will be to transition from administration on a small scale to a larger scale while maintaining the tailoring and personalization participants preferred in this study phase. Additional study challenges included intervention length. This study included 8 to 9 weeks of two-way texting between a health educator and participants. To allow SMS to serve as a true intervention booster and to increase the maintenance of study outcomes, the intervention length may need to be implemented over a longer period. We also discovered that there is one more critical persona: the health educator; that is, each participant develops a mental model of the individual with whom she would be texting. Therefore, one goal moving forward is to ensure that the health educator (for a workshop) conveys a consistent persona in future iterations of the study.

Finally, and admittedly, there is a lack of strong economic data to support the use of mHealth behavioral interventions [98,99], and this pilot study did not examine the economic implications. We will be engaging in economic evaluations in our PMI RCT feasibility pilot study (step 2 of phase 8, Table 1). We believe that the systematic design and development of mHealth solutions based on strong evidence-based foundations can begin to show economic value as critical adjuncts (PMIs) by extending the

efficacy of interventions. This requires systematically determining how a *technological form can be adapted to an evidence-based intervention.* However, this is both ill-defined and underresearched, leaving distinct gaps in the design, implementation, and acceptance of mHealth apps [100,101]. The incorporation of mHealth technologies necessitates considerations of how the current theories and practices employed can accommodate these technologies [102] and argues for the inclusion of theories and practices from other disciplines [103,104], especially those disciplines associated with explicitly studying how humans use technology [105,106]. This paper offers an approach to address such gaps.

Conclusions

In conclusion, given that digital technology is seen as an important component of addressing the STI and HIV epidemic in African American adolescents [67,107] as well as for public health innovation in general [108], we demonstrated how to systematically adapt an existing HIV and STI EBI designed for telephone-based health educator communication with African American females to one that is based on SMS texting. The ADAPT-ITT approach was used to navigate the adapted intervention. As recommended, we involved the target user group, sexual health experts, behavior change experts, software developers (SMS companies in this case), and research experts [109], a considerable strength of the study methodology. It is important to note that if mHealth technologies are to be components of an intervention, it is inappropriate to assume (explicitly or implicitly) that *technology is neutral*. This paper represents an initial step toward engaging a more systematic process, ADAPT-ITT, that will provide guidance for EBIs to extend the duration (and retain the integrity) of their impact using the most common communication technology on earth—texting.

Acknowledgments

This research was funded in part by a grant from Emory University's Global Health Institute and the Goizueta Business School's Summer Research Fund at Emory University.

Conflicts of Interest

None declared.

References

1. Koenig LJ, Hoyer D, Purcell DW, Zaza S, Mermin J. Young people and HIV: a call to action. Am J Public Health 2016 Mar;106(3):402-405. [doi: 10.2105/AJPH.2015.302979] [Medline: 26794156]
2. National HIV/AIDS Strategy For the United States: Updated to 2020 Federal Action Plan. Office of National AIDS Policy. 2015. URL: https://files.hiv.gov/s3fs-public/nhas-update.pdf [accessed 2020-09-10]
3. Monitoring Selected National HIV Prevention and Care Objectives by Using HIV Surveillance Data: United States and 6 Dependent Areas. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-25-2.pdf [accessed 2020-08-31]
4. Estimated HIV Incidence and Prevalence in the United States 2014–2018. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-25-1.pdf [accessed 2020-08-31]
5. Diagnoses of HIV Infection Among Adolescents and Young Adults in the United States and 6 Dependent Areas. Centers for Disease Control and Prevention. 2018. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2020-08-31]
6. Estimated HIV Incidence and Prevalence in the Unites States, 2014-2018. Centers for Disease Control and Prevention. 2018. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2020-08-31]
7. Kann L, McManus T, Harris W, Shanklin S, Flint K, Queen B, et al. Youth risk behavior surveillance - United States, 2017. MMWR Surveill Summ 2018 Jun 15;67(8):1-114 [FREE Full text] [doi: 10.15585/mmwr.ss6708a1] [Medline: 29902162]
8. School Health Profiles: Characteristics of Health Programs Among Secondary Schools. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/healthyyouth/data/profiles/pdf/2018/CDC-Profiles-2018.pdf [accessed 2020-08-31]
9. Mirzazadeh A, Biggs MA, Viitanen A, Horvath H, Wang LY, Dunville R, et al. Do school-based programs prevent HIV and other sexually transmitted infections in adolescents? A systematic review and meta-analysis. Prev Sci 2018 May;19(4):490-506. [doi: 10.1007/s11121-017-0830-0] [Medline: 28786046]
10. Evans R, Widman L, Stokes MN, Javidi H, Hope EC, Brasileiro J. Association of sexual health interventions with sexual health outcomes in black adolescents: a systematic review and meta-analysis. JAMA Pediatr 2020 Apr 20:- epub ahead of print. [doi: 10.1001/jamapediatrics.2020.0382] [Medline: 32310261]
11. Idele P, Gillespie A, Porth T, Suzuki C, Mahy M, Kasedde S, et al. Epidemiology of HIV and AIDS among adolescents: current status, inequities, and data gaps. J Acquir Immune Defic Syndr 2014 Jul 1;66(Suppl 2):S144-S153. [doi: 10.1097/QAI.0000000000000176] [Medline: 24918590]

12. Jenkins RA. Supplemental issue on does early intervention prevent health-risking sexual behaviors related to HIV/AIDS: commentary on effects. Prev Sci 2014 Feb;15(Suppl 1):S84-S86 [FREE Full text] [doi: 10.1007/s11121-013-0422-6] [Medline: 23881420]
13. Pettifor A, Stoner M, Pike C, Bekker L. Adolescent lives matter: preventing HIV in adolescents. Curr Opin HIV AIDS 2018 May;13(3):265-273 [FREE Full text] [doi: 10.1097/COH.0000000000000453] [Medline: 29528850]
14. Salazar LF, Bradley EL, Younge SN, Daluga NA, Crosby RA, Lang DL, et al. Applying ecological perspectives to adolescent sexual health in the United States: rhetoric or reality? Health Educ Res 2010 Aug;25(4):552-562 [FREE Full text] [doi: 10.1093/her/cyp065] [Medline: 20007196]
15. Piper K, Enah C, Daniel M. Black southern rural adolescents' HIV stigma, denial, and misconceptions and implications for HIV prevention. J Psychosoc Nurs Ment Health Serv 2014 Jun;52(6):50-56. [doi: 10.3928/02793695-20140210-01] [Medline: 24530218]
16. Murry VM, Berkel C, Chen Y, Brody GH, Gibbons FX, Gerrard M. Intervention induced changes on parenting practices, youth self-pride and sexual norms to reduce HIV-related behaviors among rural African American youths. J Youth Adolesc 2011 Sep;40(9):1147-1163 [FREE Full text] [doi: 10.1007/s10964-011-9642-x] [Medline: 21373904]
17. Buchbinder SP, Liu AY. CROI 2018: epidemic trends and advances in HIV prevention. Top Antivir Med 2018 May;26(1):1-16 [FREE Full text] [Medline: 29727292]
18. DiClemente R, Santelli J, Crosby R. Adolescent Health: Understanding and Preventing Risk Behaviors. San Francisco, CA: Jossey-Bass; 2009.
19. Lee YM, Cintron A, Kocher S. Factors related to risky sexual behaviors and effective STI/HIV and pregnancy intervention programs for African American adolescents. Public Health Nurs 2014;31(5):414-427. [doi: 10.1111/phn.12128] [Medline: 24850214]
20. Collins CB, Sapiano TN. Lessons learned from dissemination of evidence-based interventions for HIV prevention. Am J Prev Med 2016 Oct;51(4 Suppl 2):S140-S147 [FREE Full text] [doi: 10.1016/j.amepre.2016.05.017] [Medline: 27402185]
21. Thompson M. HIVMA Statement on House Labor -- HHS FY 2019 Funding bBill. HIV Medicine Association. 2018. URL: https://www.hivma.org/news_and_publications/hivma_news_releases/2018/hivma-statement-on-house-labor-hhs-fy-2019-funding-bill [accessed 2020-09-10]
22. Andrews M. The 'Perfect Storm': Redirecting Family Planning Funds Could Undercut STD Fight. Washington Post. 2018. URL: https://tinyurl.com/y6auct2f [accessed 2020-09-10]
23. Wang B, Stanton B, Deveaux L, Lunn S, Rolle G, Adderley R, et al. Multi-year school-based implementation and student outcomes of an evidence-based risk reduction intervention. Implement Sci 2017 Feb 10;12(1):16 [FREE Full text] [doi: 10.1186/s13012-016-0539-7] [Medline: 28187740]
24. Widman L, Nesi J, Kamke K, Choukas-Bradley S, Stewart JL. Technology-based interventions to reduce sexually transmitted infections and unintended pregnancy among youth. J Adolesc Health 2018 Jun;62(6):651-660 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.02.007] [Medline: 29784112]
25. Hendrick CE, Canfield C. HIV risk-reduction prevention interventions targeting African American adolescent women. Adolesc Res Rev 2017 Jun;2(2):131-149 [FREE Full text] [doi: 10.1007/s40894-016-0036-x] [Medline: 28626791]
26. Bradley E, DiClemente RJ, Sales J, Rose E, Davis T, Wingood L, et al. Make It Last: Using a Supplemental Treatment Trial Design to Maintain Effects of a Sexual Risk Reduction Intervention for African-American Adolescent Females. J Clin Trials 2014;05(02) [FREE Full text] [doi: 10.4172/2167-0870.1000210]
27. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
28. Phillips KA, Epstein DH, Mezghanni M, Vahabzadeh M, Reamer D, Agage D, et al. Smartphone delivery of mobile HIV risk reduction education. AIDS Res Treat 2013;2013:231956 [FREE Full text] [doi: 10.1155/2013/231956] [Medline: 24159383]
29. Davis TL, DiClemente R, Prietula M. Taking mhealth forward: examining the core characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]
30. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
31. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
32. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
33. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered ehealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]

34. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
35. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
36. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
37. Radovic A, Badawy SM. Technology use for adolescent health and wellness. Pediatrics 2020 May;145(Suppl 2):S186-S194. [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
38. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
39. Chernick LS, Berrigan M, Gonzalez A, Konja A, Stockwell MS, Ehrhardt A, et al. Engaging adolescents with sexual health messaging: a qualitative analysis. J Adolesc Health 2019 Nov;65(5):660-666. [doi: 10.1016/j.jadohealth.2019.05.029] [Medline: 31495641]
40. Madden M, Lenhart A, Duggan M, Cortesi S, Gasser U. Teens and Technology. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/03/13/teens-and-technology-2013/ [accessed 2020-08-31]
41. Perry RC, Kayekjian KC, Braun RA, Cantu M, Sheoran B, Chung PJ. Adolescents' perspectives on the use of a text messaging service for preventive sexual health promotion. J Adolesc Health 2012 Sep;51(3):220-225. [doi: 10.1016/j.jadohealth.2011.11.012] [Medline: 22921131]
42. Smith A. African Americans and Technology Use: A Demographic Portrait. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/2014/01/06/african-americans-and-technology-use/ [accessed 2020-08-31]
43. Gilliam M, Chor J, Hill B. Digital media and sexually transmitted infections. Curr Opin Obstet Gynecol 2014 Oct;26(5):381-385. [doi: 10.1097/GCO.0000000000000104] [Medline: 25105560]
44. Rempel GR, Ballantyne RT, Magill-Evans J, Nicholas DB, Mackie AS. Texting teens in transition: the use of text messages in clinical intervention research. JMIR Mhealth Uhealth 2014 Nov 6;2(4):e45 [FREE Full text] [doi: 10.2196/mhealth.3232] [Medline: 25379624]
45. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
46. Rideout V, Robb M. Social Media, Social Life: Teens Reveal Their Experiences. Common Sense Media.: Common Sense Media; 2018. URL: https://www.commonsensemedia.org/research/social-media-social-life-2018 [accessed 2020-09-10]
47. Anderson M, Jiang J. Teens, Social Media & Technology. Public Services Alliance. 2018. URL: http://publicservicesalliance.org/wp-content/uploads/2018/06/Teens-Social-Media-Technology-2018-PEW.pdf [accessed 2020-08-31]
48. Walther JB. Interpersonal effects in computer-mediated interaction. Commun Res 2016 Jun 30;19(1):52-90. [doi: 10.1177/009365092019001003]
49. Walther JB. Computer-mediated communication. Commun Res 2016 Jun 29;23(1):3-43. [doi: 10.1177/009365096023001001]
50. Subrahmanyam K, Smahel D, Greenfield P. Connecting developmental constructions to the internet: identity presentation and sexual exploration in online teen chat rooms. Dev Psychol 2006 May;42(3):395-406. [doi: 10.1037/0012-1649.42.3.395] [Medline: 16756432]
51. Subrahmanyam K, Reich SM, Waechter N, Espinoza G. Online and offline social networks: use of social networking sites by emerging adults. J Appl Dev Psychol 2008 Nov;29(6):420-433. [doi: 10.1016/j.appdev.2008.07.003]
52. Dunbar R, Arnaboldi V, Conti M, Passarella A. The structure of online social networks mirrors those in the offline world. Soc Netw 2015 Oct;43:39-47. [doi: 10.1016/j.socnet.2015.04.005]
53. Ehrenreich SE, Beron KJ, Burnell K, Meter DJ, Underwood MK. How adolescents use text messaging through their high school years. J Res Adolesc 2020 Jun;30(2):521-540. [doi: 10.1111/jora.12541] [Medline: 31868974]
54. Lenhart A, Smith A, Anderson M, Duggan M, Perrin A. Teens, Technology & Friendships. Pew Research Center. 2015. URL: https://www.pewresearch.org/wp-content/uploads/sites/9/2015/08/Teens-and-Friendships-FINAL2.pdf [accessed 2020-08-31]
55. Song JH, Hollenbeck CR. The value of social presence in mobile communications. Serv Ind J 2015 Jul 10;35(11-12):611-632. [doi: 10.1080/02642069.2015.1062880]
56. Cornelius JB, Dmochowski J, Boyer C, St Lawrence J, Lightfoot M, Moore M. Text-messaging-enhanced HIV intervention for African American adolescents: a feasibility study. J Assoc Nurses AIDS Care 2013;24(3):256-267 [FREE Full text] [doi: 10.1016/j.jana.2012.06.005] [Medline: 23122907]
57. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]

58. Oinas-Kukkonen H. A foundation for the study of behavior change support systems. Pers Ubiquit Comput 2012 Jul 19;17(6):1223-1235. [doi: 10.1007/s00779-012-0591-5]
59. Fogg B. A Behavior Model for Persuasive Design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive'09; April 26-29, 2009; Claremont , USA. [doi: 10.1145/1541948.1541999]
60. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
61. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
62. Coomes CM, Lewis MA, Uhrig JD, Furberg RD, Harris JL, Bann CM. Beyond reminders: a conceptual framework for using short message service to promote prevention and improve healthcare quality and clinical outcomes for people living with HIV. AIDS Care 2012;24(3):348-357. [doi: 10.1080/09540121.2011.608421] [Medline: 21933036]
63. Wingood GM, DiClemente RJ. The ADAPT-ITT model: a novel method of adapting evidence-based HIV Interventions. J Acquir Immune Defic Syndr 2008 Mar 1;47(Suppl 1):S40-S46. [doi: 10.1097/QAI.0b013e3181605df1] [Medline: 18301133]
64. Chambers DA, Norton WE. The adaptome: advancing the science of intervention adaptation. Am J Prev Med 2016 Oct;51(4 Suppl 2):S124-S131 [FREE Full text] [doi: 10.1016/j.amepre.2016.05.011] [Medline: 27371105]
65. Escoffery C, Lebow-Skelley E, Udelson H, Böing EA, Wood R, Fernandez ME, et al. A scoping study of frameworks for adapting public health evidence-based interventions. Transl Behav Med 2019 Jan 1;9(1):1-10 [FREE Full text] [doi: 10.1093/tbm/ibx067] [Medline: 29346635]
66. McKleroy VS, Galbraith JS, Cummings B, Jones P, Harshbarger C, Collins C, ADAPT Team. Adapting evidence-based behavioral interventions for new settings and target populations. AIDS Educ Prev 2006 Aug;18(4 Suppl A):59-73. [doi: 10.1521/aeap.2006.18.supp.59] [Medline: 16987089]
67. DiClemente RJ, Bradley E, Davis TL, Brown JL, Ukuku M, Sales JM, et al. Adoption and implementation of a computer-delivered HIV/STD risk-reduction intervention for African American adolescent females seeking services at county health departments: implementation optimization is urgently needed. J Acquir Immune Defic Syndr 2013 Jun 1;63 Suppl 1:S66-S71 [FREE Full text] [doi: 10.1097/QAI.0b013e318292014f] [Medline: 23673891]
68. Wingood GM, Simpson-Robinson L, Braxton ND, Raiford JL. Design of a faith-based HIV intervention: successful collaboration between a university and a church. Health Promot Pract 2011 Nov;12(6):823-831. [doi: 10.1177/1524839910372039] [Medline: 21511996]
69. Latham TP, Sales JM, Boyce LS, Renfro TL, Wingood GM, DiClemente RJ, et al. Application of ADAPT-ITT: adapting an evidence-based HIV prevention intervention for incarcerated African American adolescent females. Health Promot Pract 2010 May;11(3 Suppl):53S-60S. [doi: 10.1177/1524839910361433] [Medline: 20488969]
70. Sullivan PS, Stephenson R, Grazter B, Wingood G, Diclemente R, Allen S, et al. Adaptation of the African couples HIV testing and counseling model for men who have sex with men in the United States: an application of the ADAPT-ITT framework. Springerplus 2014;3:249 [FREE Full text] [doi: 10.1186/2193-1801-3-249] [Medline: 24877036]
71. Khumsaen N, Stephenson R. Adaptation of the HIV/AIDS self-management education program for men who have sex with men in Thailand: an application of the ADAPT-ITT framework. AIDS Educ Prev 2017 Oct;29(5):401-417 [FREE Full text] [doi: 10.1521/aeap.2017.29.5.401] [Medline: 29068714]
72. Myers B, Carney T, Browne FA, Wechsberg WM. Development of a trauma-informed substance use and sexual risk reduction intervention for young South African women. Patient Prefer Adherence 2018;12:1997-2006 [FREE Full text] [doi: 10.2147/PPA.S175852] [Medline: 30323569]
73. Davidson TM, Lopez CM, Saulson R, Borkman AL, Soltis K, Ruggiero KJ, et al. Development and preliminary evaluation of a behavioural HIV-prevention programme for teenage girls of Latino descent in the USA. Cult Health Sex 2014;16(5):533-546 [FREE Full text] [doi: 10.1080/13691058.2014.891049] [Medline: 24697607]
74. Shrestha R, Altice F, Karki P, Copenhaver M. Developing an integrated, brief biobehavioral HIV prevention intervention for high-risk drug users in treatment: the process and outcome of formative research. Front Immunol 2017;8:561 [FREE Full text] [doi: 10.3389/fimmu.2017.00561] [Medline: 28553295]
75. Bradley EL, Sutton MY, Cooks E, Washington-Ball B, Gaul Z, Gaskins S, et al. Developing FAITHH: methods to develop a faith-based HIV stigma-reduction intervention in the rural south. Health Promot Pract 2018 Sep;19(5):730-740 [FREE Full text] [doi: 10.1177/1524839917754044] [Medline: 29383967]
76. Takishima-Lacasa JY, Kameoka VA. Adapting a sexually transmitted infection prevention intervention among female adolescents in Hawai'i. Health Promot Pract 2019 Jul;20(4):608-615. [doi: 10.1177/1524839918769592] [Medline: 29644890]
77. Cavanaugh CE, Campbell J, Braxton N, Harvey J, Wingood G. Adapting an evidence-based HIV-prevention intervention for women in domestic violence shelters. Psychol Violence 2016 Jul;6(3):469-477 [FREE Full text] [doi: 10.1037/vio0000042] [Medline: 27398257]
78. Wingood GM, Reddy P, Lang DL, Saleh-Onoya D, Braxton N, Sifunda S, et al. Efficacy of SISTA South Africa on sexual behavior and relationship control among isiXhosa women in South Africa: results of a randomized-controlled trial. J Acquir

Immune Defic Syndr 2013 Jun 1;63(Suppl 1):S59-S65 [FREE Full text] [doi: 10.1097/QAI.0b013e31829202c4] [Medline: 23673889]
79. DiClemente RJ, Wingood GM, Sales JM, Brown JL, Rose ES, Davis TL, et al. Efficacy of a telephone-delivered sexually transmitted infection/human immunodeficiency virus prevention maintenance intervention for adolescents: a randomized clinical trial. JAMA Pediatr 2014 Oct;168(10):938-946 [FREE Full text] [doi: 10.1001/jamapediatrics.2014.1436] [Medline: 25155070]
80. Compendium of Evidence-Based Interventions and Best Practices for HIV Prevention. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/research/interventionresearch/compendium/rr/index.html [accessed 2020-08-31]
81. Bandura A. Social cognitive theory and exercise of control over HIV infection. In: DiClemente R, Peterson J, editors. Preventing AIDS: Theories and Methods of Behavioral Interventions. New York, USA: Plenum Press; 1994.
82. Gibbons MC. Gender and Power: Society, The Person and Sexual Politics. Stanford, CA: Stanford University Press; 1987.
83. Wingood GM, DiClemente RJ. Application of the theory of gender and power to examine HIV-related exposures, risk factors, and effective interventions for women. Health Educ Behav 2000 Oct;27(5):539-565. [doi: 10.1177/109019810002700502] [Medline: 11009126]
84. Wingood G, Camp C, Dunkle K, Cooper H, DiClemente R. The theory of gender and power: constructs, variables, and implications for developing HIV interventions for women. In: DiClemente R, Crosby R, Kegler M, editors. Emerging Theories in Health Promotion Practice and Research. San Francisco, CA: Wiley; 2009.
85. Wingood GM, DiClemente RJ. Enhancing adoption of evidence-based HIV interventions: promotion of a suite of HIV prevention interventions for African American women. AIDS Educ Prev 2006 Aug;18(4 Suppl A):161-170. [doi: 10.1521/aeap.2006.18.supp.161] [Medline: 16987097]
86. Lenhart A, Smith A, Anderson M. Teens, Technology and Romantic Relations. Pew Research Center. 2015. URL: http://www.pewinternet.org/files/2015/04/PI_TeensandTech_Update2015_0409151.pdf [accessed 2020-08-31]
87. Smith A. Americans and Text Messaging. Pew Research Center. 2011. URL: https://www.pewresearch.org/internet/2011/09/19/americans-and-text-messaging/ [accessed 2020-08-31]
88. Kuniavsky M. Observing the User Experience: A Practitioner's Guide to User Research. Waltham, MA: Morgan Kaufmann; 2012.
89. Tullis T, Albert B. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Waltham, MA: Morgan Kaufmann; 2013.
90. Hartson R, Pyla P. The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Waltham, MA: Morgan Kaufmann; 2012.
91. Jeminiwa RN, Hohmann NS, Fox BI. Developing a theoretical framework for evaluating the quality of mhealth apps for adolescent users: a systematic review. J Pediatr Pharmacol Ther 2019;24(4):254-269 [FREE Full text] [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
92. Mabey D. Epidemiology of sexually transmitted infections: worldwide. Medicine 2014 Jun;42(6):287-290. [doi: 10.1016/j.mpmed.2014.03.004]
93. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
94. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
95. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
96. Glasgow RE, Eckstein ET, Elzarrad MK. Implementation science perspectives and opportunities for HIV/AIDS research: integrating science, practice, and policy. J Acquir Immune Defic Syndr 2013 Jun 1;63(Suppl 1):S26-S31. [doi: 10.1097/QAI.0b013e3182920286] [Medline: 23673882]
97. Kelton K, Fleischmann KR, Wallace WA. Trust in digital information. J Am Soc Inf Sci 2008 Feb 1;59(3):363-374. [doi: 10.1002/asi.20722]
98. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
99. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
100. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]

101. van Velsen L, Wentzel J, van Gemert-Pijnen JE. Designing ehealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
102. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
103. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
104. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
105. Poole ES. HCI and mobile health interventions: how human-computer interaction can contribute to successful mobile health interventions. Transl Behav Med 2013 Dec;3(4):402-405 [FREE Full text] [doi: 10.1007/s13142-013-0214-3] [Medline: 24294328]
106. Buller DB, Berwick M, Shane J, Kane I, Lantz K, Buller MK. User-centered development of a smart phone mobile application delivering personalized real-time advice on sun protection. Transl Behav Med 2013 Sep;3(3):326-334 [FREE Full text] [doi: 10.1007/s13142-013-0208-1] [Medline: 24058385]
107. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas L, Wingood G, et al. A computational future for preventing HIV in minority communities: how advanced technology can improve implementation of effective programs. J Acquir Immune Defic Syndr 2013 Jun 1;63(Suppl 1):S72-S84 [FREE Full text] [doi: 10.1097/QAI.0b013e31829372bd] [Medline: 23673892]
108. DiClemente R, Nowara A, Shelton R, Wingood G. Need for innovation in public health research. Am J Public Health 2019 Feb;109(S2):S117-S120. [doi: 10.2105/AJPH.2018.304876] [Medline: 30785791]
109. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C. Sexual health promotion for young people delivered via digital media: a scoping review. Public Health Res 2015;3(13). [doi: 10.3310/phr03130] [Medline: 26583166]

Abbreviations

ADAPT-ITT: assessment, decision, adaptation, production, topical experts-integration, training, testing
CASI: computer-assisted self-interview
CDC: Centers for Disease Control and Prevention
EBI: evidence-based intervention
GRA: graduate research assistant
mHealth: mobile health
PMI: preventive maintenance intervention
RCT: randomized controlled trial
STI: sexually transmitted infection
TGP: theory of gender power
UX: user experience

Edited by G Eysenbach; submitted 15.07.20; peer-reviewed by S Badawy; comments to author 06.08.20; revised version received 18.08.20; accepted 23.08.20; published 06.10.20.

Please cite as:
Davis T, DiClemente RJ, Prietula M
Using ADAPT-ITT to Modify a Telephone-Based HIV Prevention Intervention for SMS Delivery: Formative Study
JMIR Form Res 2020;4(10):e22485
URL: https://formative.jmir.org/2020/10/e22485
doi:10.2196/22485
PMID:32831178

Original Paper

A Web-Based, Mobile-Responsive Application to Screen Health Care Workers for COVID-19 Symptoms: Rapid Design, Deployment, and Usage

Haipeng Zhang[1,2,3], DO, MMSc; Dimitar Dimitrov[4], MSSE; Lynn Simpson[4], MPH; Nina Plaks[5], MS; Balaji Singh[5], BSc; Stephen Penney[5], MBA; Jo Charles[6], MPH; Rosemary Sheehan[7], MBA; Steven Flammini[5], BS; Shawn Murphy[4,8], MD, PhD; Adam Landman[1,5,9], MD, MS, MIS, MHS

[1]Digital Innovation Hub, Brigham and Women's Hospital, Boston, MA, United States
[2]Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA, United States
[3]Department of Medicine, Harvard Medical School, Boston, MA, United States
[4]Research Information Science and Computing, Partners HealthCare, Boston, MA, United States
[5]Partners Information Systems, Partners HealthCare, Somerville, MA, United States
[6]Partners Enterprise Data and Digital Health, Partners HealthCare, Boston, MA, United States
[7]Partners Human Resources, Partners HealthCare, Somerville, MA, United States
[8]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States
[9]Department of Emergency Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Haipeng Zhang, DO, MMSc
Digital Innovation Hub
Brigham and Women's Hospital
75 Francis St
Boston, MA, 02115
United States
Phone: 1 6174591650
Email: hzhang37@partners.org

Abstract

Background: As of July 17, 2020, the COVID-19 pandemic has affected over 14 million people worldwide, with over 3.68 million cases in the United States. As the number of COVID-19 cases increased in Massachusetts, the Massachusetts Department of Public Health mandated that all health care workers be screened for symptoms daily prior to entering any hospital or health care facility. We rapidly created a digital COVID-19 symptom screening tool to enable this screening for a large, academic, integrated health care delivery system, Partners HealthCare, in Boston, Massachusetts.

Objective: The aim of this study is to describe the design and development of the COVID Pass COVID-19 symptom screening application and report aggregate usage data from the first three months of its use across the organization.

Methods: Using agile principles, we designed, tested, and implemented a solution over the span of one week using progressively customized development approaches as the requirements and use case become more solidified. We developed the minimum viable product (MVP) of a mobile-responsive, web-based, self-service application using research electronic data capture (REDCap). For employees without access to a computer or mobile device to use the self-service application, we established a manual process where in-person, socially distanced screeners asked employees entering the site if they have symptoms and then manually recorded the responses in an Office 365 Form. A custom .NET Framework application solution was developed as COVID Pass was scaled. We collected log data from the .NET application, REDCap, and Microsoft Office 365 from the first three months of enterprise deployment (March 30 to June 30, 2020). Aggregate descriptive statistics, including overall employee attestations by day and site, employee attestations by application method (COVID Pass automatic screening vs manual screening), employee attestations by time of day, and percentage of employees reporting COVID-19 symptoms, were obtained.

Results: We rapidly created the MVP and gradually deployed it across the hospitals in our organization. By the end of the first week, the screening application was being used by over 25,000 employees each weekday. After three months, 2,169,406 attestations were recorded with COVID Pass. Over this period, 1865/160,159 employees (1.2%) reported positive symptoms. 1,976,379 of

the 2,169,406 attestations (91.1%) were generated from the self-service screening application. The remainder were generated either from manual attestation processes (174,865/2,169,406, 8.1%) or COVID Pass kiosks (25,133/2,169,406, 1.2%). Hospital staff continued to work 24 hours per day, with staff attestations peaking around shift changes between 7 and 8 AM, 2 and 3 PM, 4 and 6 PM, and 11 PM and midnight.

Conclusions: Using rapid, agile development, we quickly created and deployed a dedicated employee attestation application that gained widespread adoption and use within our health system. Further, we identified 1865 symptomatic employees who otherwise may have come to work, potentially putting others at risk. We share the story of our implementation, lessons learned, and source code (via GitHub) for other institutions who may want to implement similar solutions.

(JMIR Form Res 2020;4(10):e19533) doi:10.2196/19533

KEYWORDS

public health; clinical informatics; digital health; coronavirus; COVID-19; SARS-CoV-2; 2019-nCov; app; eHealth

Introduction

To date, over 14 million cases of COVID-19 have been confirmed worldwide, with over 3.68 million cases in the United States [1]. This number continues to grow, and the United States has become the epicenter of COVID-19 [2]. By April 5, 2020, in the Commonwealth of Massachusetts, there were already 12,500 confirmed cases of COVID-19 and 231 deaths [3]. With exponentially increasing numbers of COVID-19 cases, the need for digital technology to address issues arising from pandemics such as COVID-19 has grown considerably [4].

To limit the spread and "flatten the curve," on March 16, 2020, the Massachusetts Department of Public Health (MDPH) and the Commissioner of Public Health issued an order that Massachusetts hospitals must screen all visitors, including employees, for symptoms of a respiratory infection (fever, cough, shortness of breath, or sore throat) and that individuals with any symptoms should not be permitted to visit the hospitals [5]. Shortly after, our institution enacted a policy that all employees working in a patient care facility must wear a face mask while working as another measure to limit the spread of COVID-19 within the health care workforce [6].

We created a digital symptom screening and attestation tool, called COVID Pass, that provides daily facility passes and face mask passes for employees as its output. We used a prototype-driven innovation model combined with a transition to a more traditional custom development team as the application requirements matured, which allowed our group to rapidly deploy and refine the solution as it was released and approached scale. In the first 2 weeks, COVID Pass transitioned from a paper proposal to an enterprise-supported solution used by over 25,000 employees daily. After 3 months, over 2 million attestations were performed using the application.

In this paper, we describe the design, development, and use of the COVID Pass application, and we make the code available for other institutions seeking to implement a similar solution.

Methods

This study was conducted at Partners HealthCare, a not-for-profit, academic, integrated health care delivery system in Boston, Massachusetts. Partners include Brigham and Women's Hospital, Massachusetts General Hospital, community and specialty hospitals, a physician network, community health centers, home care, a health insurance plan, and other health-related services. The largest private employer in Massachusetts, Partners HealthCare has approximately 74,000 employees, including physicians, nurses, scientists, and caregivers.

Complying with the MDPH requirement for symptom screening in such a large, widely geographically distributed organization was expected to be difficult. The Partners HealthCare Chief Human Resources Officer recognized an opportunity to link the distribution of masks (something that employees wanted) with completion of the symptom screening (something that might be more difficult to have all employees complete). The Human Resources team and Occupational Health Services provided the initial requirements, including an application that would enable employees who must work on-site at a facility that provides direct patient care to be able to self-screen for symptoms of COVID-19 infection prior to being allowed to enter the facility. This application needed to be mobile-responsive, provide guidance to the employee about next steps if they indicated symptoms were present, create a pass that would be "glanceable" to entrance way screening staff, and be capable of exporting user logs on at least a daily basis. A manual pathway where on-site screeners would ask employees about symptoms and record them was also created to address the needs of employees who were unable to use the electronic self-screening tool.

Using agile principles, we designed, tested, and implemented a solution over the span of a week using progressively customized development approaches as the requirements and use case become more solidified. Based on the requirements, we developed the minimum viable product (MVP) of the self-service application using research electronic data capture (REDCap) due to the speed at which we could develop a functional prototype using the prebuilt data export systems of this solution. REDCap is a secure, web-based software platform that is designed to support data capture for research studies, providing an intuitive interface for validated data capture; audit trails for tracking data manipulation and export procedures; automated export procedures for seamless data downloads to common statistical packages; and procedures for data integration and interoperability with external sources [7,8].

The self-service application, initially developed in REDCap, was called COVID Pass. The application required users to log

in with a Partners HealthCare network user ID and password. The application then authenticated the user against Active Directory and used the login ID to look up the employee's first name, last name, email address, and employee ID. The employee would review this information and select whether they had symptoms of COVID-19 from a list of symptoms determined by infection control leaders (Figure 1). Some symptoms were marked as NEW in the system to better differentiate chronic symptoms from acute onset symptoms. If the employee selected "no symptoms," they were required to attest to this with their initials and were then provided a pass to enter the facility for the day. The pass was displayed on the screen, and a copy was automatically sent to the employee's email address (Figure 2). Employees indicating one or more symptoms were directed not to come to work and to call their manager and Occupational Health Services (Figure 3). The occupational health team would receive a daily report of positive symptom attestations and follow up according to their standard protocols. A copy of the relevant information was also sent to the employee's email address. Figure 4 shows a schematic of the processes and procedures involved in using the self-service component of COVID Pass.

Figure 1. Employee symptom reporting screen of the COVID Pass application.

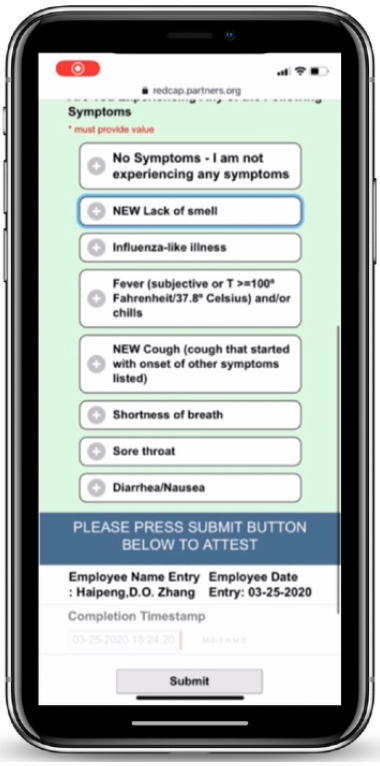


Figure 2. "Cleared for Work" screen of the COVID Pass application.

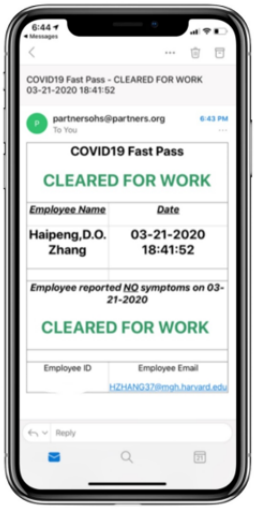
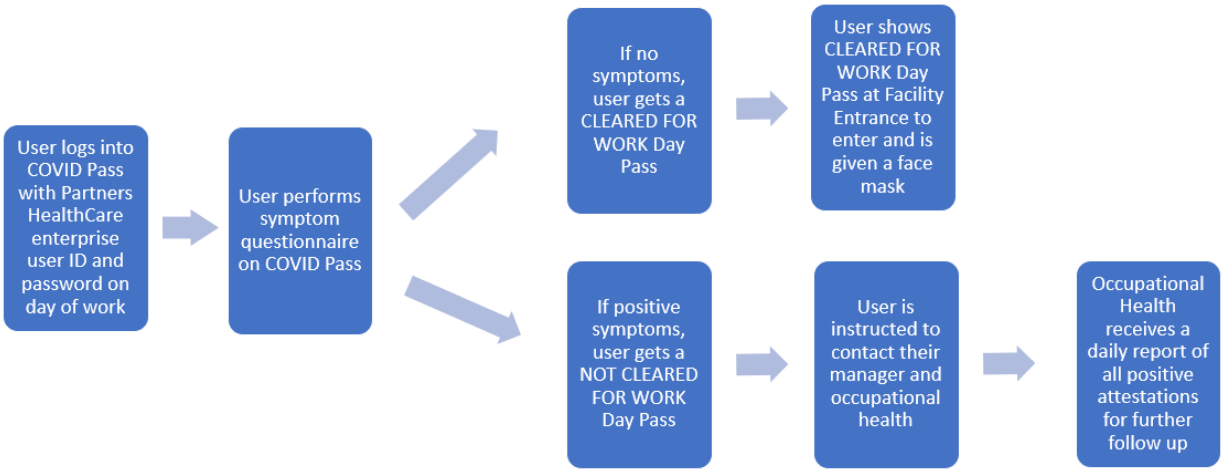


Figure 3. "Not Cleared for Work" screen of the COVID Pass application.

Figure 4. Schematic of the process and procedures involved in using the COVID Pass application.



We made COVID Pass available to employees via multiple methods. Working with marketing and

internal communications experts, we selected an easy-to-remember URL,[9]. We also created a quick response (QR) code for the URL that that users could scan with their smartphones to be directed to the website. Finally, we wrapped the website in native iOS and Android apps so that the application could be distributed internally through our employee App Catalog.

For employees who did not have access to a computer or mobile device to complete COVID Pass, a manual process was established where in-person, socially distanced screeners asked employees entering the site if they had symptoms and then manually recorded the responses. For the manual pathway MVP, we created custom Microsoft Office 365 forms (Microsoft Corporation) that enabled screeners at each of the pilot sites to quickly record the employee's first and last name as well as whether they had symptoms.

A custom .NET Framework application (Microsoft Corporation) was developed by a separate development team as the REDCap version of COVID Pass was being refined and its requirements solidified (Figure 5). In this version, both the self-service attestation mode and the manual process were rolled into one build. In addition, a kiosk mode version of COVID Pass was created for employees to provide attestations at the entrances of facilities. As COVID Pass scaled, we ultimately transitioned from the REDCap application to the custom .NET application on March 30, 2020, at 5 PM.

Figure 5. Screenshot of the .NET Framework version of COVID Pass.

All versions of COVID Pass recorded employee attestations and stored them in one central database. The data were aggregated at the site level and shared daily with site operational leaders; names of individuals reporting positive symptoms were shared with Human Resources leadership daily. In this report, we share three months of data from COVID Pass use from March 30, 2020, to June 30, 2020.

Aggregate descriptive statistics, including overall employee attestations by day and site, employee attestations by application method (COVID Pass automatic screening vs manual screening vs kiosk), employee attestations by time of day, and percentage of employees reporting COVID-19 symptoms, were compiled using SAS (Enterprise Guide 7.1, SAS Institute Inc) and Tableau (Desktop 2020.2, Tableau Software LLC). This study was reviewed by the Partners HealthCare Institutional Review Board and deemed to not meet the definition of human subjects research.

The COVID Pass MVP was initially tested at two hospitals within our system on March 23, 2020. As part of this controlled rollout, the project team participated in on-site testing of the application, gathering user feedback and being "at the elbow" with the staff at the entranceways of the sites. Figure 6 shows an image of a hospital entrance with a COVID Pass lane for employees who completed COVID Pass prior to or immediately upon arrival. The manual screening lane includes a table at which screeners can enter the employees' attestations manually. Feedback from both staff assigned to screen employees and employees using COVID Pass was relayed to the developer for further refinement of the application. In parallel to this development process, the number of sites within the system using COVID Pass was slowly increased throughout the week. By March 30, 2020, COVID Pass was deployed across the enterprise.

Figure 6. Hospital staff entrance showing the COVID Pass lane for employees who used the COVID Pass application and the manual screening lane for employees who did not use the application.

Documentation about the initiation of COVID Pass within a facility was collated and distributed to the sites as new COVID Pass sites went live. This documentation included standard language, collateral such as flyers used in previous site implementations, and "lessons learned" from earlier implementations. The collective documentation also simplified the process of scaling this solution across the enterprise.

Results

Multimedia Appendix 1 summarizes the number of employee attestations (using either COVID Pass, the manual process, or kiosks) by week and site. 1,976,379 of the 2,169,387 total attestations (91.1%) were generated from the self-service screening application. The remainder (Table 1) were generated either from manual attestation processes (174,865/2,169,387, 8.1%) or COVID Pass kiosks (25,133/2,169,387, 1.2%). Hospital staff continued to work 24 hours per day, with peak staff attestations during shift changes between 6 and 8 AM, 1 and 3 PM, 5 and 7 PM, and 10 PM and midnight (Figure 7).

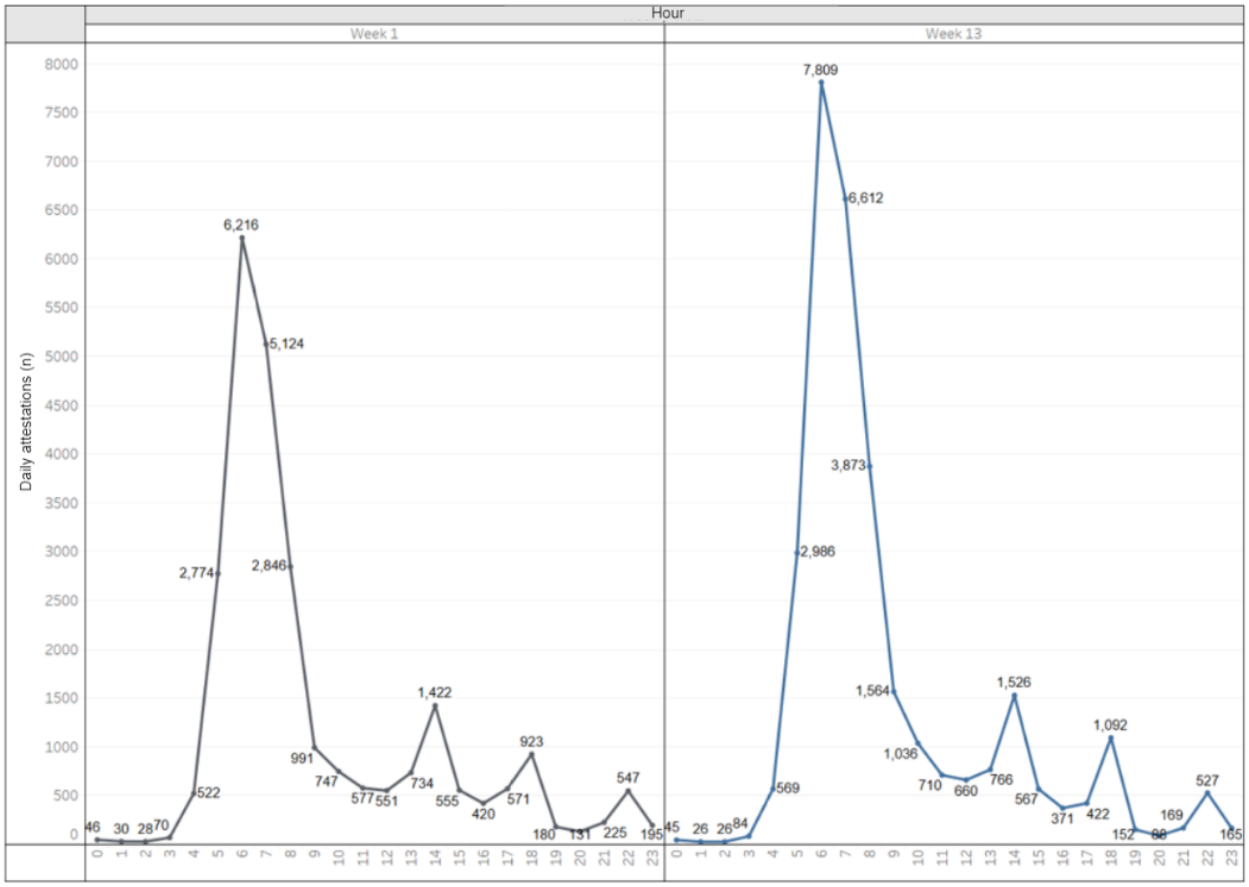
As the number of COVID-19 cases stabilized in the region, more employees transitioned back to working on-site. Figure 7 presents a comparison between the average daily (Monday to Friday) number of hourly attestations in week 1 (week of March 30) vs week 13 (week of June 22). By week 13, the average number of attestations at the peak 6 AM shift change had increased by 1593 (25.63%).

There was an overall increase of 26,179 attestations weekly by week 13 compared to week 1, representing a 14.46% increase in attestations.

Table 1. COVID Pass attestations by employee app-based self-attestation, manual screening, and kiosk by site on March 30 to June 30, 2020 (N=2,169,387). Note that some sites used separate manual screening data collection methods; data for these sites are not reflected here.

Site number (total attestations, n)	Attestation source, n (%)		
	App	Kiosk	Manual
1 (732,278)	713,259 (97.4)	1552 (0.2)	17,467 (2.4)
2 (564,710)	497,101 (88.0)	16,129 (2.9)	51,480 (9.1)
3 (166,985)	140,964 (88.4)	2466 (1.5)	23,555 (14.1)
4 (148,918)	127,010 (85.3)	133 (0.1)	21,775 (14.6)
5 (107,191)	104,368 (97.4)	0 (0.0)	2823 (2.6)
6 (93,142)	70,124 (75.3)	4763 (5.1)	18,255 (10.6)
7 (66,581)	64,838 (97.4)	18 (<0.1)	1725 (2.6)
8 (57,773)	51,260 (88.7)	0 (0.0)	6513 (11.3)
9 (50,552)	34,845 (68.9)	0 (0.0)	15,707 (37.9)
10 (47,532)	44,650 (93.9)	4 (<0.1)	2878 (6.1)
11 (32,766)	32,080 (97.9)	22 (0.1)	664 (2.0)
12 (32,176)	24,010 (74.6)	0 (0.0)	8166 (23.4)
13 (25,744)	24,934 (96.9)	0 (0.0)	810 (3.1)
14 (15,713)	13,438 (85.5)	1 (<0.1)	2274 (14.5)
15 (14,132)	13,654 (96.6)	0 (0.0)	478 (3.4)
16 (8437)	8400 (99.6)	1 (<0.1)	36 (0.4)
17 (1837)	1803 (98.1)	31 (1.7)	3 (0.2)
18 (1048)	896 (85.5)	0 (0.0)	152 (14.5)
19 (658)	568 (86.3)	6 (0.9)	84 (12.8)
20 (643)	643 (100.0)	0 (0.0)	0 (0.0)
21 (571)	534 (93.5)	7 (1.2)	30 (5.3)
All sites (2,169,387)	1,969,379 (90.8)	25,133 (1.2)	174,875 (8.1)

Figure 7. Average daily employee attestations from Monday to Friday (COVID Pass, manual screening, and kiosk) by hour of day during Week 1 (March 30 to April 3, 2020) and Week 13 (June 22 to 26, 2020).



Discussion

Principal Findings

Within two weeks of conception, COVID Pass went live across the entire Partners HealthCare enterprise. During weekdays, COVID Pass recorded over 25,000 attestations per day. This high rate of adoption would not have occurred at such a fast pace without the following key considerations: mandatory and incentivized use while minimizing friction; planning for accessibility; leveraging REDCap to enable an agile development process; rapid analysis and distribution of COVID Pass data; and pre-emptive transition to a more durable and scalable platform.

By making COVID Pass mandatory for employees to gain access to their work sites and incentivizing it by pairing it with the distribution of masks, we quickly increased the acceptability of the potential burden of mandatory prework self-attestation. We believed that COVID Pass would further benefit from minimization of friction during completion of the attestation process. We achieved this by minimizing the number of interactions users needed to perform to complete the COVID Pass process while maximizing the data captured. By autopopulating demographic information after login, we only required the user to answer one question (symptom review). If the user was asymptomatic, we required them to attest to this with their initials to receive their COVID Pass for the day. We also simplified access to COVID Pass through the creation of multiple distribution channels, including a simple URL that was sent with all communication to staff throughout our organization, a QR code that was used on flyers in the entranceways of our facilities, and iOS and Android app versions of COVID Pass that were made available through our employee-facing App Catalog.

As part of this work, we also knew that we would need to provide an accessible pathway for users who might not be able to complete the self-service COVID Pass due to limited proficiency with a smartphone or computer, language limitations, or other reasons. By incorporating this accessibility requirement early in our development process and creating a manual pathway, we ensured that COVID Pass would be a comprehensive solution for all employees. Eventually, we also added a kiosk mode for COVID Pass, along with multilingual support to further enhance accessibility.

We adopted an agile development process to further minimize friction for the end user through early on-site testing and rapid iteration cycles. An agile development approach emphasizes “early and continuous delivery of valuable software” while accommodating changing requirements as the software is used in the real world and assumptions are validated or invalidated [10]. Our team created the MVP for COVID Pass within 48 hours and began testing the solution the very next day. We refined the requirements for COVID Pass and made updates to the MVP multiple times throughout the day at the beginning of this process.

Much of this early ability to create rapid changes and adjustments to COVID Pass came from the initial platform decision to use REDCap for the MVP. As the core functionality of REDCap was able to accommodate most of the initial requirements of COVID Pass, the symptom survey fields and conditional text could be adjusted almost instantaneously. This enabled our team to sustain a rapid iteration cycle during the first week of deployment while additional sites across the organization went live with COVID Pass.

In addition, REDCap enabled our team to rapidly export data and custom reports from COVID Pass to our organization daily. These data were primarily used to perform two major functions within the organization: (1) tracking symptomatic employees for further workup of COVID-19 by occupational health and (2) estimating the number of employees on site for personal protective equipment (PPE) supply planning.

Within the first week, COVID Pass had identified over 500 employees with symptoms suggestive of COVID-19 infection. After three months, COVID Pass had identified 1865 employees with symptoms. These employees were all instructed to contact Occupational Health Services for next steps. In addition, all employees with positive symptoms were flagged to be contacted by Occupational Health Services for further questioning and triage. Further work is needed to determine how many of these employees were actually confirmed to have COVID-19.

The reproduction number of COVID-19 has been preliminarily estimated to be between 2 and 4 [11,12]. By identifying 1865 employees with symptoms suggestive of COVID-19 and instructing them to not report to work, and by alerting Occupational Health Services about these employees for further follow-up, COVID Pass likely played an important role in limiting further spread of COVID-19 within our workforce, our patients, and the larger community in our region.

The World Health Organization estimates that up to 89 million medical masks are required monthly for the COVID-19 response [13]. As all COVID Pass users who are cleared for work are issued a face mask, the daily logs from COVID Pass have become an important proxy for the number of masks being distributed daily to employees within our organization. This also provides leaders within the organization with a good approach to predict face mask allocations per facility per day and plan accordingly.

By the middle of the first week of the COVID Pass rollout, it became apparent that a more hardened version of this application would be needed. Some of the clear advantages of REDCap, such as the prebuilt scaffolding to support survey creation and data export, became limitations as more custom change requests began to surface for COVID Pass. In addition, as COVID Pass was rapidly adopted across the enterprise, a more formalized support structure was needed to maintain the operation of COVID Pass 24 hours per day, 7 days per week. As a result, we worked in parallel with a second, in-house development team to create a custom version of COVID Pass built on the Microsoft .NET Framework with a structured query language (SQL) database, using the lessons learned and updated requirements gained through the release of the live REDCap application.

We ultimately transitioned to the .NET version of COVID Pass on March 30, 2020, at 5 PM. The process of developing the same solution in parallel and on two different platforms allowed us to leverage the strengths of each one and achieve a more robust final product. The established processes and large user base for COVID Pass meant that a compiled, single-purpose application, such as the .NET application, was better suited for the long-term needs of the project. Transitioning the application to an infrastructure environment that provided 24/7 support and available resource capacity enabled the system to handle high load times during the morning hours, when the highest number of concurrent users occurs. A refined self-service module and a manual pathway built on the same platform enabled streamlining of the data collection and analysis as well.

Limitations

We custom-developed COVID Pass to meet local requirements for employee symptom attestation. Other organizations may have different operational requirements. We are making both the REDCap and .NET source code available to other organizations to use and modify as needed. Secondly, some of our sites did not track manual attestation using the Office 365 forms or .NET applications; therefore, we do not possess complete data on manual screenings. Finally, as the scope of this paper is focused on the data generated directly from the COVID Pass applications, we have no data on the number of positive or negative COVID-19 tests given to employees who were flagged by COVID Pass to Occupational Health Services.

Conclusion

Health systems worldwide have faced incredible challenges due to the COVID-19 pandemic. At Partners HealthCare, one challenge was the evolving role of symptom monitoring for all employees working on-site. Technology is not a panacea; however, when it is used appropriately and scoped to the right problems, technology can play a meaningful role in the era of COVID-19.

With the combination of a rapid, agile approach to software development and a use case paired with an end-user incentive, we quickly created and deployed a dedicated employee attestation application that gained widespread adoption and use within our health system. COVID Pass is continuing to support daily screening for over 30,000 employees. Further, we identified 1865 symptomatic employees who otherwise may have come to work, potentially putting others at risk.

We continue to use the data obtained through COVID Pass for incident planning, such as PPE supply. The COVID Pass platform also has potential to be an effective communication tool to the workforce; one use case that we are exploring is using the final page of COVID Pass to inform employees about voluntary COVID-19 research studies within our institution.

As we share these lessons learned and the story of our implementation of COVID Pass, we also want to ensure that other institutions that may want to implement a similar solution can not only learn from our implementation but can also access the source code for COVID Pass. Therefore, we are making the source code for COVID Pass available to all via GitHub [14].

We are living through extraordinary times. We are confident that the bravery, commitment, and perseverance of the clinicians and hospital staff on the front lines of this worldwide crisis will help see us safely through the COVID-19 pandemic. Our hope is that software solutions such as COVID Pass will play their own small roles within this larger effort and that others may make meaningful use of this tool, as our institution has.

Conflicts of Interest

ABL is a consultant for the Abbott Medical Device Cybersecurity Council.

Multimedia Appendix 1
Weekly completed employee symptom attestations (COVID Pass and manual) by day and site.
[DOCX File , 19 KB - formative_v4i10e19533_app1.docx]

References

1. COVID Tracker. Google. URL: https://www.google.com/search?q=covid+tracker&rlz=1C1GCEA_enUS874US875&oq=covid+tracker&aqs=chrome.69i57j0l7.3341j1j7&sourceid=chrome&ie=UTF-8 [accessed 2020-07-18]
2. COVID-19 Dashboard. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-04-05]
3. Coronavirus Disease 2019 (COVID-19) Cases in MA. Massachusetts Department of Public Health. 2020 Apr 05. URL: https://www.mass.gov/doc/covid-19-cases-in-massachusetts-as-of-april-5-2020/download [accessed 2020-04-05]
4. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/nejmp2005835]
5. Hospital Visitor Restrictions Guidance March 16 2020. Mass.gov. 2020 Mar 16. URL: https://www.mass.gov/doc/hospital-visitor-restrictions-guidance-march-16-2020 [accessed 2020-04-05]
6. Fox JC. All Partners HealthCare employees now required to wear masks while on duty. Boston Globe. 2020 Mar 22. URL: https://www.bostonglobe.com/2020/03/22/metro/all-mgh-employees-now-required-wear-masks-while-duty/ [accessed 2020-04-05]
7. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
8. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
9. COVID Pass. Partners HealthCare. 2020. URL: https://www.partners.org/covidpass [accessed 2020-09-09]
10. 12 Principles Behind the Agile Manifesto. Agile Alliance. 2015. URL: https://www.agilealliance.org/agile101/12-principles-behind-the-agile-manifesto/ [accessed 2020-04-06]
11. Zhao S, Lin Q, Ran J, Musa S, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020 Mar;92:214-217 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.050] [Medline: 32007643]
12. Riou J, Althaus C. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro Surveill 2020 Jan;25(4):30 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.4.2000058] [Medline: 32019669]
13. Shortage of personal protective equipment endangering health workers worldwide. World Health Organization. 2020 Mar 03. URL: https://www.who.int/news-room/detail/03-03-2020-shortage-of-personal-protective-equipment-endangering-health-workers-worldwide [accessed 2020-04-09]
14. Partners HealthCare. GitHub. URL: https://github.com/partnershealthcare [accessed 2020-09-09]

Abbreviations

MDPH: Massachusetts Department of Public Health
MVP: minimum viable product
PPE: personal protective equipment
QR: quick response
REDCap: research electronic data capture
SQL: structured query language

Edited by G Eysenbach; submitted 22.04.20; peer-reviewed by Y Chu, DA Banerjee, P Deluca; comments to author 06.06.20; revised version received 18.07.20; accepted 11.08.20; published 08.10.20.

Please cite as:
Zhang H, Dimitrov D, Simpson L, Plaks N, Singh B, Penney S, Charles J, Sheehan R, Flammini S, Murphy S, Landman A
A Web-Based, Mobile-Responsive Application to Screen Health Care Workers for COVID-19 Symptoms: Rapid Design, Deployment, and Usage
JMIR Form Res 2020;4(10):e19533
URL: https://formative.jmir.org/2020/10/e19533
doi:10.2196/19533
PMID:32877348

Original Paper

Integration of Online Treatment Into the "New Normal" in Mental Health Care in Post–COVID-19 Times: Exploratory Qualitative Study

Joyce J P A Bierbooms[1,2], PhD; Monique van Haaren[2], MSc; Wijnand A IJsselsteijn[3], Prof Dr; Yvonne A W de Kort[3], Prof Dr; Milou Feijt[3], MSc; Inge M B Bongers[1,2], Prof Dr

[1]Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

[2]Mental Healthcare Eindhoven, Eindhoven, Netherlands

[3]Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Joyce J P A Bierbooms, PhD
Tranzo
Tilburg School of Social and Behavioral Sciences
Tilburg University
PO Box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 630642496
Email: J.J.P.A.Bierbooms@tilburguniversity.edu

Abstract

Background: The COVID-19 pandemic has necessitated an immediate and large-scale uptake of online treatment for mental health care. However, there is uncertainty about what the "new normal" in mental health care will be like in post–COVID-19 times. To what extent will the experiences gained during the pandemic influence a sustainable adoption and implementation of online mental health care treatment in the future?

Objective: In this paper, we aim to formulate expectations with regard to the sustainability of online mental health care after COVID-19.

Methods: In an interview study, 11 mental health care professionals were asked about their experiences and expectations for the future. Participants were recruited from a mental health care organization in the Netherlands. The interviews took place between April 7-30, 2020, at the peak of the COVID-19 crisis in the Netherlands. The data were analyzed using a thematic coding method.

Results: From the interviews, we learn that the new normal in mental health care will most likely consist of more blended treatments. Due to skill enhancement and (unexpected) positive experiences with online treatment, an increase in adoption is likely to take place. However, not all experiences promise a successful and sustainable upscaling of online treatment in the future. Mental health care professionals are learning that not all clients are able to benefit from this type of treatment.

Conclusions: Sustainable upscaling of online mental health care requires customized solutions, investments in technology, and flexibility of mental health care providers. Online treatment could work for those who are open to it, but many factors influence whether it will work in specific situations. There is work to be done before online treatment is inherently part of mental health care.

(JMIR Form Res 2020;4(10):e21344) doi:10.2196/21344

KEYWORDS

online treatment; sustainability; mental health care; COVID-19

Introduction

In recent months, the COVID-19 pandemic and subsequent governmental regulations in the Netherlands have urged mental health care providers to accommodate an immediate and large-scale uptake of online treatment. Up to that point, many mental health care professionals were still quite hesitant, despite the proven benefits, to use online treatment options [1,2]. Previous studies have shown that the adoption of online treatment is influenced by a variety of factors, such as a lack of digital skills among both practitioners and clients, technical issues, the assumption that technology-mediated treatment may not allow for real interpersonal contact, and (lack of) prior experiences regarding the potential added value of online treatment [3-5].

Many questions currently are being raised about what the "new normal" of society will constitute, if and when COVID-19 has run its course. This also applies to mental health care. Both optimism and concerns have been expressed about the acceptance of online treatment as a new normal in mental health care [6-10]. There is optimism about improving the accessibility of mental health care by upscaling online treatment and accommodating clients to receive therapy in a way that easily fits into their daily lives. This may empower clients, foster their self-efficacy, and enable them to engage in their treatment (more) independently of time and place (eg, [8,10]). However, concerns are expressed that not everyone will benefit equally from this game changer. Specifically, vulnerable mental health clients may lack the digital skills, cognitive ability, motivation, and/or resources to partake successfully in online digital treatment, and are at risk of being disconnected from the care they need (eg, [7,9]).

Online treatment as part of the new normal in mental health care will—and must—depend on mental health care professionals' and clients' experiences regarding their use of online treatment during the COVID-19 pandemic [10]. These experiences are therefore important sources of feedback that will help anticipate the effects of this period of "forced upscaling" on the sustainability of online treatment in mental health care and the new normal that we will find ourselves in.

Based on mental health care professionals' experiences during the COVID-19 pandemic, the current paper seeks to discuss the expectations regarding the sustainability of online treatment in mental health care when it is no longer necessitated by COVID-19 regulations.

Methods

Design

We used an exploratory qualitative study design in which we interviewed mental health care professionals about their experiences with online treatment during the COVID-19 period and their expectations regarding the new ways of working in mental health care when this crisis is controlled.

Context

In the Netherlands, mental health care is provided within a system of frontline services by general practitioners; primary and secondary mental health care is provided by specialized mental health care providers. Funding depends on the type of treatment that is required and can be provided by the national government (long-term secondary mental health care), health insurance companies (primary and short-term secondary mental health care), or the local government (youth mental health care). Online treatment is seen as an important development in mental health care and is stimulated accordingly. For example, while waiting lists can be long in mental health care, health insurance companies have made arrangements with mental health care professionals to provide online treatment to people on the waiting list. This is fully funded although people are not yet registered as clients [11].

This study was conducted at Stichting Geestelijke Gezondheidszorg Eindhoven en de Kempen (GGzE), a mental health care provider in Eindhoven, the Netherlands. GGzE has approximately 2200 employees (including management and supporting staff) and provides clinical and ambulatory mental health care to more than 20,000 clients. Mental health care is provided either in group therapy or individually and can be face to face or online. In 2019, GGzE started an online treatment team providing full online treatment trajectories (ie, using videoconferencing, online modules, and messaging services). This online treatment team currently has a caseload of approximately 200 clients with light to complex psychological and psychiatric problems.

Participants

For the interviews, we recruited 9 mental health care professionals and 2 eHealth support staff employees from GGzE. In the sample of mental health care professionals, 5 had recently started working online, and 4 were already experienced, having used online treatment tools for more than 1 year. The latter group comprised those who specifically chose to be an online therapist and were therefore already positive about it before they started, in contrast to the therapists who were recently "forced" to start with online treatment since face-to-face therapy was not possible. During the COVID-19 pandemic, online treatment services consisted of videoconferencing, online treatment modules, and messaging services using a secured communication platform. Whereas a broad range of other online treatment tools are available (eg, virtual reality, self-monitoring apps), the upscaling of online treatment in the context of the COVID-19 crisis primarily concerns the aforementioned tools. During the COVID-19 period, an average of 1060 therapy sessions were conducted online per week versus 59 online consultations per week in the first 6 weeks of 2020.

The participants were approached after obtaining consent from their manager. They were provided with an invitational email in which the purpose of the interview was explained; then volunteers were contacted by the researcher. The purpose of the interview was explained once again, and the participants were asked to sign an informed consent form.

Data Collection

Data collection took place between April 7-30, 2020, when the COVID-19 crisis was at its peak and mental health care professionals had been working online between 3-6 weeks. Mental health care professionals who were new to using online treatment were asked about their experiences related to performing treatment online, their skill improvement, and their intentions to use online treatment in the future. Mental health care professionals who had been using online treatments for >1 year were asked about the barriers they had encountered prior to becoming an online therapist, their current experiences, and their strategies for establishing rapport with their clients. The supporting staff was asked which questions and concerns they encountered while supporting professionals with online treatment tool use. All participants were asked about their expectations regarding the continued use of online psychological treatment in post–COVID-19 times.

Data Analysis

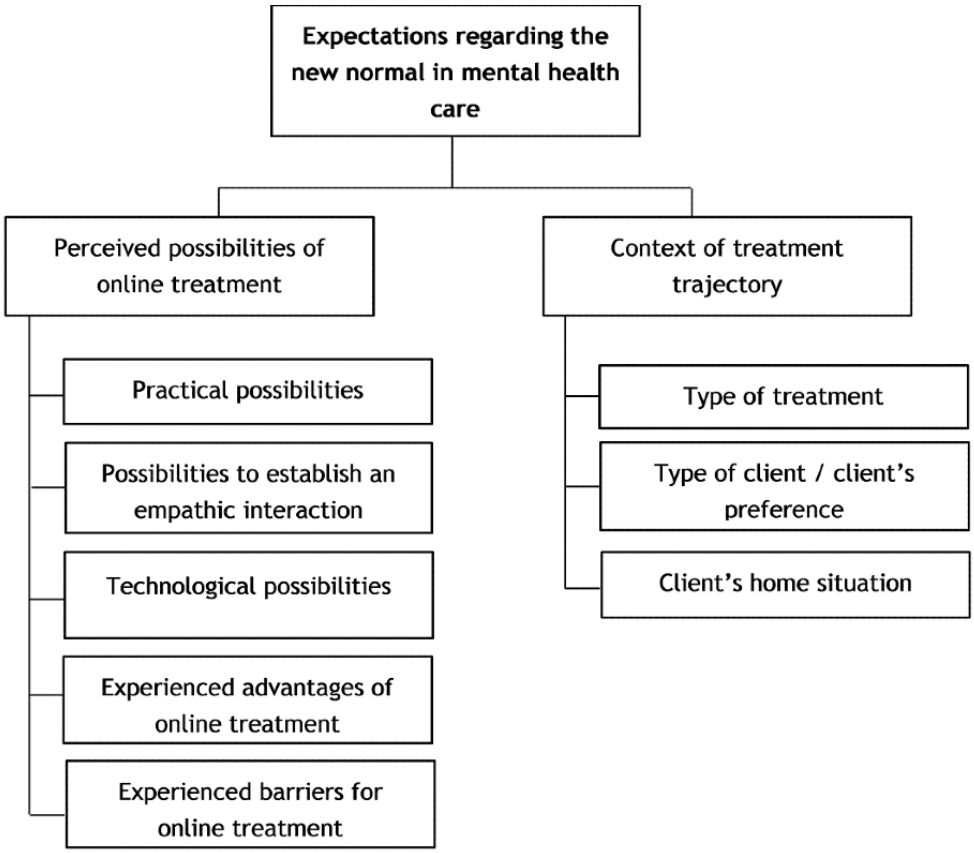
The interviews were recorded and transcribed verbatim. The data were analyzed using a thematic coding approach [12]. Themes were identified based on the purpose of the study. Subsequently, all fragments of the data were coded and assigned to subthemes.

Results

Overview

From the data analysis, two main themes could be derived regarding the role of online treatment in the new normal in mental health care: the perceived possibilities of online treatment and the context of a treatment trajectory. The data revealed several subthemes that represent the issues that contribute to this new normal based on current experiences during the COVID-19 period from the perspective of the mental health care professionals themselves and how they perceive their clients to experience online treatment. Figure 1 illustrates the main themes and subthemes.

Figure 1. Themes and subthemes.



The results of the interviews provide preliminary insight into what the new normal in mental health care and the role of online treatment could look like. There were a number of (unexpected) positive experiences, but there were also concerns. On the one hand, the results suggest that because important barriers such as a lack of skills and experience have now been lowered, online treatment will be used more frequently. On the other hand, some professionals also expect that they, or their colleagues, will revert to regular face-to-face treatment as soon as possible, since a number of professionals and clients perceive regular face-to-face treatment as the only way to create a therapeutic alliance. In general, we found that the participants in this study who were already experienced online practitioners were more confident about the continuation of online treatment in the future compared to others.

Perceived Possibilities of Online Treatment

Overall, the results show that mental health care professionals come to understand that much more is possible in online treatment than what they had initially expected, for example,

drawing something on paper or sharing one's screen to go through an online module together with a client. Mental health care professionals also report that their clients tend to be more open when they are in their home environment and that genuine empathic contact is indeed possible in remote communication. These positive effects surprised the mental health care professionals who started delivering online treatment during the COVID-19 period; the experienced online practitioners were already convinced of these possibilities. One respondent predicted that because clients are now experiencing the benefits of online treatment, there is a significant chance that they will continue to ask for it, which will to a large extent determine whether online treatment options will be integrated into the new normal. Furthermore, mental health care professionals expect that they will more frequently consider online treatment as an alternative now that they are aware of its potential value. Reported examples of such added value include the reduction in travel time, an increase in self-efficacy and activation in clients, more openness by clients when they are in their home environment, and increased accessibility to treatment due to flexibilities in time and place. This, in turn, leads to more frequent, shorter consultations, which adds to a higher level of connectedness between client and therapist.

> *I do this by writing a message each morning asking my client how he slept and then I can give some suggestions to get through the day. [...] I noticed, and this is surprising to me, that this delivers epistemic trust from my client towards me.* [P1]

> *In a face-to-face meeting, you tend to fill the 45 minutes you planned, despite the fact that you only need 20 minutes, because your client came all the way over to your office. When I have an online meeting with my client, I can much more easily finish the meeting after 20 minutes when we have nothing else to talk about anymore.* [P4]

The future position of online treatment in mental health care will also be determined by the technological conditions required to conduct online treatments. Mental health care professionals report having experienced high levels of frustration over technical issues they encountered while trying to get their online environments to work. These problems are felt by both new and experienced online therapists. In addition, clients do not always possess the necessary technical devices or a robust internet connection, nor do they always have an appropriate and quiet place with sufficient privacy where they can talk to their therapist. There is also the issue of confidence in the security of the system, which in some cases withholds professionals, but mainly clients, from wanting to perform therapy online. Moreover, some clients receive therapy without their family/relatives' knowledge, which makes it difficult to engage in online sessions from home.

> *Especially now that the kids are at home it is very difficult for clients to find a quiet place to talk to me. Sometimes they are calling by telephone from the car because they could not find a proper place at home.* [P6]

A struggle that is also mentioned, mostly by the more unexperienced online therapists, is that it is very difficult to connect at an interpersonal level during an online interaction, such as videoconferencing. Nonverbal cues can be easily missed, which requires a number of additional questions and explicit feedback moments from a therapist. Moreover, participants stated multiple times that it can be very difficult to react adequately to clients' emotions in an online therapeutic setting.

> *It can be very difficult to react to a client's emotions, for example, when someone starts to cry, I find it difficult to react when I am not in the same room. There is literally a distance.* [P9]

Context of a Treatment Trajectory

According to mental health care professionals, the continued use of online treatment in the future will depend on the type of treatment and the type of clients—there will not be a one-size-fits-all solution. On this issue, both experienced and unexperienced online mental health care practitioners agree. The resilience that is now shown by a number of clients in responding to online treatment does not apply to every one of them. According to the participants, not all clients are equally susceptible to online treatments. The general experience is that group therapy does not work well online, nor do therapy sessions with multiple participants from the social network of a client. In addition, Eye Movement Desensitization and Reprocessing (EMDR) is considered close to impossible to perform online by some professionals. Interestingly, others report unexpected positive experiences regarding complex therapies such as EMDR and imagery rescripting. For those with positive experiences with these complex therapies, the benefits that were mentioned earlier (eg, reduction of travel time) will probably determine whether these forms of therapy will be continued remotely after the COVID-19 period. There is also no consensus regarding the necessity to do at least the intakes in a face-to-face setting. In particular, therapists who are more experienced in online treatment tend to claim they have very effective online intakes for various complex mental health care problems. This may imply that the number of online intakes might increase to at least some extent in comparison to the pre–COVID-19 situation, due to experiences gained in this period.

> *Some colleagues argue that an intake cannot be done online, but I have very positive experiences with online intakes.* [P7]

Discussion

Overview

As the COVID-19 mitigation measures are slowly being lifted in the Netherlands, mental health care organizations have started to think about expanding possibilities beyond regular face-to-face treatment. Numerous pleas, opinions, and discussions regarding the continuation of online treatment in mental health care are being prompted in professional literature, on professional network platforms, and on social media. The question we raised at the beginning of this paper becomes increasingly urgent with each step mental health care takes in

liberalizing COVID-19 measures: what new normal will we adopt regarding online treatment in mental health care?

Expectations

Based on therapists' experiences with online treatment during the COVID-19 period, we can create a clearer image of the possibilities of online treatment in mental health care and how a sustainable increase in the use of online tools can be reached. As it is clear that experience and skill enhancement will lead to fewer barriers among mental health care professionals to use online treatment tools [13-15], there will likely be more online treatment in the near future than there was before COVID-19. We also argue that some of the positive experiences will convince even therapists who were previously resistant toward online treatment to consider an online session when this saves time and is more convenient for the client. A recent study shows that, in most cases, the perceived advantages regarding online treatment outweigh the disadvantages [3]. However, we also have reason to believe that in some cases online therapy may not work. There are issues that are more difficult to overcome, such as clients having an inadequate environment without proper (technological) facilities to receive online treatment or clients for whom it is difficult to open up to this concept. The conviction that a real therapeutic alliance can be achieved through remote communication does not seem to come simply by gaining experience in using online treatment. Additionally, frequent users of online treatment and therapists who are generally positive about the possibilities prefer to see people face to face in certain situations. The bottom line is that mental health care delivery in the new normal will most likely result in more blended forms of care. The sustainability of blended care requires giving space to diversity and being flexible in providing online treatment for those who benefit from it, while providing alternatives for those who do not.

Limitations

This paper describes a small study with 11 participants from one particular mental health care organization. This means that further research is needed to strengthen the results. Despite the small sample size, data saturation was reached and many mental health care professionals also indicated that were disclosing ideas and experiences of their colleagues. Another limitation is that clients were not interviewed in this round. This was due to the limited amount of time in which we wanted to draw a preliminary overview of experiences with online treatment in mental health care. A focus on mental health care professionals was chosen because of the large impact the adoption among therapists has on a sustainable implementation of online treatment. The clients' experiences that are described in this paper are reflections of practitioners based on their interactions with their clients about online treatment.

Future Research

The sustainability of online treatment in mental health care will depend on the complex interplay of individual, social, organizational, and economic factors [16]. The exact implications of the COVID-19 crisis for online mental health care need further exploration taking into account the complexity of the situation. In our future work, we will investigate the sustainability of online treatment in mental health care in post–COVID-19 time in relation to the significant changes the pandemic and its aftermath will continue to impose upon mental health care. We will continue to work with mental health care organizations in addressing the challenges they face in effectively delivering online and blended forms of treatment, including skill enhancement in online treatment of mental health care professionals and investigating the acceptance and use of online treatment tools by both therapists and clients. In addressing these challenges, our focus is on innovative solutions aimed to optimize the accessibility, effectiveness, efficiency, and treatment quality of online treatment tools, including their various technological and user requirements as well as the organizational conditions that encourage a sustainable uptake of online treatment.

Acknowledgments

This research was carried out as part of a VIPP3 GGZ Koplopers 2019-2021 grant project, which is funded by the Dutch Ministry of Health, Welfare and Sport (grant no. 329644). The authors would like to thank the mental health care professionals who were willing to take the time to share their experiences with us. The COVID-19 pandemic has imposed large pressure on both mental health care professionals and their clients. We truly appreciate the time taken in these exigent circumstances to participate in this interview study.

Conflicts of Interest

None declared.

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;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
2. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
3. Connolly S, Miller C, Lindsay J, Bauer M. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clin Psychol Sci Pract 2020 Jan 06;27(2):e12311 [FREE Full text] [doi: 10.1111/cpsp.12311]

4. 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 Dec 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
5. 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]
6. Basu T. The coronavirus pandemic is a game changer for mental health care. MIT Technology Review. 2020 Mar 20. URL: https://www.technologyreview.com/2020/03/20/905184/coronavirus-online-therapy-mental-health-app-teletherapy/ [accessed 2020-05-18]
7. Spoelstra S. Van face to face naar zorg op afstand: De ggz na de coronacrisis. Zorgvisie 2020;50(4):20-23 [FREE Full text] [doi: 10.1007/s41187-020-0367-x] [Medline: 32421051]
8. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the Crisis Into an Opportunity: Digital Health Strategies Deployed During the COVID-19 Outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
9. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
10. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: The 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]
11. CZ zet Welshop in voor online behandeling GGz (CZ uses Welshop for online treatment in mental healthcare). 2019 Apr 17. URL: https://www.icthealth.nl/nieuws/cz-zet-welshop-in-voor-online-behandeling-ggz/ [accessed 2020-08-20]
12. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
13. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
14. Donovan C, Poole C, Boyes N, Redgate J, March S. Australian mental health worker attitudes towards cCBT: What is the role of knowledge? Are there differences? Can we change them? Internet Interventions 2015 Nov;2(4):372-381 [FREE Full text] [doi: 10.1016/j.invent.2015.09.001] [Medline: 32595044]
15. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental Health Care Goes Online: Practitioners' Experiences of Providing Mental Health Care During the COVID-19 Pandemic. Cyberpsychol Behav Soc Netw 2020 Aug 18. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
16. Wensing M, Grol R, Grimshaw J, editors. Improving Patient Care: The Implementation of Change in Health Care, Third Edition. Chichester, West Sussex: John Wiley & Sons Ltd; 2020.

Abbreviations

EMDR: Eye Movement Desensitization and Reprocessing
GGzE: Stichting Geestelijke Gezondheidszorg Eindhoven en de Kempen

Edited by J Torous; submitted 12.06.20; peer-reviewed by L Ebenfeld, N Rauseo-Ricupero; comments to author 30.07.20; revised version received 24.08.20; accepted 05.09.20; published 08.10.20.

Please cite as:
Bierbooms JJPA, van Haaren M, IJsselsteijn WA, de Kort YAW, Feijt M, Bongers IMB
Integration of Online Treatment Into the "New Normal" in Mental Health Care in Post–COVID-19 Times: Exploratory Qualitative Study
JMIR Form Res 2020;4(10):e21344
URL: http://formative.jmir.org/2020/10/e21344/
doi:10.2196/21344
PMID:33001835

Original Paper

Understanding Problems With Sleep, Sexual Functioning, Energy, and Appetite Among Patients Who Access Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy for Anxiety and Depression: Qualitative Exploratory Study

Michael R Edmonds[1*], MA; Heather D Hadjistavropoulos[1*], PhD; Kirsten M Gullickson[1*], PhD; Aleiia JN Asmundson[1*]; Blake F Dear[2*], PhD; Nickolai Titov[2*], PhD

[1]Online Therapy Unit, Department of Psychology, University of Regina, Regina, SK, Canada

[2]eCentreClinic, Department of Psychology, Macquarie University, Sydney, Australia

[*]all authors contributed equally

Corresponding Author:
Heather D Hadjistavropoulos, PhD
Online Therapy Unit
Department of Psychology
University of Regina
3737 Wascana Pkwy
Regina, SK
Canada
Phone: 1 306 585 5133
Email: hadjista@uregina.ca

Abstract

Background: Transdiagnostic internet-delivered cognitive behavioral therapy (T-ICBT) is an effective treatment for anxiety and depression, and nowadays, there is interest in exploring ways to optimize T-ICBT in routine care. T-ICBT programs are designed to address the primary cognitive-affective and behavioral symptoms of anxiety and depression (eg, low mood, worry, anhedonia, and avoidance). Treatment also has the potential to resolve other symptom concerns (eg, sleep disruption, sexual dysfunction, lack of energy, and appetite or weight changes). Having additional information regarding the extent of these concerns and how concerns change over time could prove beneficial for further development of T-ICBT in routine care.

Objective: This exploratory formative study aims to better understand sleep, sexual functioning, energy, and appetite concerns among T-ICBT clients seeking treatment for depression and anxiety. A qualitative analytic approach was used to identify themes in the symptom concerns reported by patients in the areas of sleep, sexual functioning, energy, and appetite at the time of enrollment. Patient responses to related items from screening measures for anxiety and depression were also examined pre- and posttreatment.

Methods: Patients in routine care who applied for a T-ICBT program for depression and anxiety over a 1-year period were included in this study. As part of the application and screening process, participants completed depression and anxiety symptom measures (ie, 9-item Patient Health Questionnaire and 7-item Generalized Anxiety Disorder scale). These same measures were administered posttreatment. Subsequently, they were asked if they were experiencing any problems with sleep, sexual activity, energy, or appetite (yes or no). If their response was yes, they were presented with an open-ended comment box that asked them to describe the problems they had experienced in those areas.

Results: A total of 462 patients were admitted to T-ICBT during the study period, of which 438 endorsed having some problems with sleep, sexual activity, energy, or appetite. The analysis of open-ended responses indicated that 73.4% (339/462) of patients reported sleep problems (eg, difficulty initiating or maintaining sleep), 69.3% (320/462) of patients reported problems with energy or motivation (eg, tiredness and low motivation), 57.4% (265/462) of patients reported appetite or body weight concerns (eg, changes in appetite and weight loss or gain), and 30.1% (139/462) of patients described concerns with sexual functioning (eg, loss of interest in sex and difficulty with arousal). Item analysis of symptom measures demonstrated that T-ICBT produced improvements in sleep, energy, and appetite in 8 weeks. Sexual dysfunction and weight changes were not represented in the screening measures, so it remains unclear what effect T-ICBT has on these symptoms.

Conclusions: Sleep disruption, lack of energy, appetite or weight changes, and sexual dysfunction are common concerns reported by clients enrolled in T-ICBT in routine practice and may deserve greater attention in T-ICBT program development and administration.

(JMIR Form Res 2020;4(10):e15037) doi:10.2196/15037

KEYWORDS

cognitive behavioral therapy; anxiety; depression; internet-based intervention; sleep; sexual health; motivation; appetite

Introduction

Background

Anxiety and depression are prevalent, disabling, and frequently comorbid mental health conditions that are associated with a range of negative outcomes, including a number of physical health problems, lower marital satisfaction, absenteeism or presenteeism in the workplace, and financial difficulties [1]. Transdiagnostic internet-delivered cognitive behavioral therapy (T-ICBT) has emerged as an efficient and effective treatment for individuals who experience concurrent symptoms of anxiety and depression [2]. T-ICBT generally involves clients completing weekly web-based lessons and homework that teach cognitive and behavioral techniques for managing symptoms [3]. The transdiagnostic approach is based on the premise that anxiety and depression are often comorbid and share common symptoms, such as maladaptive thinking patterns and avoidance [4,5]. T-ICBT protocols typically teach clients a variety of symptom management strategies that are applicable to both anxiety and depression, including cognitive restructuring, behavioral activation, and graded exposure [6]. Although some T-ICBT programs are unguided (ie, the patient completes the program independently), many are therapist assisted (ie, the patient has regular contact with a web-based therapist). In therapist-assisted T-ICBT programs, support is typically provided through secure messaging or over the phone wherein therapists answer patient questions, offer guidance and encouragement, and provide feedback on symptom changes [3].

Considerable evidence supports the efficacy of T-ICBT, and as a result, it is increasingly being implemented in routine care [7]. T-ICBT protocols are particularly appealing because they streamline the treatment process for patients and therapists [2,4]. From the patient's perspective, T-ICBT produces equally impressive treatment effects as disorder-specific internet-delivered cognitive behavioral therapy (ICBT) while saving patients from the burden of having to complete several treatment protocols sequentially to address comorbid conditions [2,4,8-11]. From the therapist's perspective, T-ICBT protocols simplify referral decision making and training for staff who can learn and implement one transdiagnostic program rather than several disorder-specific programs [2,4]. Although T-ICBT programs show considerable promise, it remains a relatively new form of treatment in routine care; thus, it is important to learn from patients and continually seek to improve treatment to meet the needs of patients [12].

T-ICBT programs are designed to address the primary cognitive-affective and behavioral symptoms of anxiety and depression (eg, low mood, worry, anhedonia, and avoidance), yet successful treatment has the potential to resolve many additional symptom concerns that are present with anxiety and depression. For example, internalizing disorders such as anxiety and depression share common features, such as sleep problems, sexual dysfunction, loss of energy, and changes in appetite/weight [13-20]; however, little is known about T-ICBT clients' concerns in these areas at the beginning of T-ICBT and the impact of T-ICBT on these additional concerns. Formative research that seeks to improve understanding of diverse symptoms pre- and posttreatment could be beneficial for identifying opportunities to improve the delivery of T-ICBT in routine care.

Objectives

The goal of this exploratory study is two-fold. First, we aim to better understand the nature of client concerns about sleep, sexual functioning, energy, and appetite/weight at the time of enrollment by analyzing responses to an open-ended question about problems with sleep, sexual functioning, energy, and appetite or weight at pretreatment. Qualitative research is an excellent method for examining patient concerns [21]. Second, following an approach similar to qualitative assurance studies [22], item-level data from screening measures of anxiety and depression were examined to explore whether T-ICBT was associated with symptom improvements in these areas at posttreatment. Given the nature of depression and anxiety, we hypothesized that clients would commonly report experiencing problems with sleep, sexual functioning, energy, or appetite/weight. In terms of the nature of the clients' concerns, we expected clients to report sleep problems such as insomnia or hypersomnia, sexual problems such as reduced sexual desire and difficulties with arousal, fatigue or tiredness, and changes in appetite and accompanying weight change. We also hypothesized that item analysis would show improvements in these areas posttreatment; however, we did not have hypotheses about the magnitude of symptom reduction.

Methods

Study Design and Ethics

This study involved extracting patient records from the Online Therapy Unit, which is a government-funded web-based therapy clinic at the University of Regina that freely offers ICBT on a routine basis to residents of Saskatchewan. In addition to providing T-ICBT, the Online Therapy Unit conducts research to optimize the delivery of ICBT in routine care; thus, patients who receive T-ICBT are asked for consent to use their deidentified data for research purposes. This study involved retrieving records from 462 patients who received T-ICBT with once-a-week therapist support over a 1-year period (June 1, 2016, to May 31, 2017). We did not retrieve additional data

from patients who received ICBT for chronic pain or received T-ICBT without regular weekly therapist support. Of note, 245 of the 462 patient records were part of a registered trial (ISRCTN14230906) [23], but as they received standard T-ICBT with the same once-a-week therapist support, their data were included in this study.

Participants and Procedure

Individuals who participated in this study reported that they heard about the Online Therapy Unit through a variety of sources, including referral from a community mental health clinic (170/462, 36.5%), referral by their family physician (101/462, 21.7%), word of mouth (69/462, 14.8%), media (57/462, 12.2%), searches on the web and email announcements (55/462, 11.8%), or printed posters or cards (14/462, 3%). Individuals who wished to take part in the Wellbeing Course began by completing a brief web-based screening questionnaire. The questionnaire ensured that participants met the basic inclusion criteria, including (1) being aged 18 years or older; (2) residing in Saskatchewan, Canada; (3) endorsing symptoms of depression and/or anxiety; (4) being comfortable using a computer with access to the internet; and (5) willing to consent to treatment and provide a medical contact for emergency purposes. Of note, a formal diagnosis of an anxiety or mood disorder was not required to participate in the program, but all patients endorsed at least some symptoms of anxiety or depression.

Those who met the basic inclusion criteria then completed additional web-based screening questions about their background (eg, demographic characteristics and contact information) and symptoms (eg, anxiety and depression). Specifically, patients completed the 7-item Generalized Anxiety Disorder (GAD-7) scale and the 9-item Patient Health Questionnaire (PHQ-9), widely used screening measures that ask patients to report the frequency of anxiety and depression symptoms over the previous 2 weeks [24,25]. The PHQ-9 and GAD-7 are formatted and scaled very similarly, with response options ranging from 0 (*not at all*) to 3 (*nearly every day*), resulting in possible scores between 0 and 21 for the GAD-7 and between 0 and 27 for the PHQ-9. Of relevance to this study were items 3, 4, and 5 from the PHQ-9, which assessed symptoms of sleep, energy, and weight. Participants with scores greater than 5 on the GAD-7 or PHQ-9 were considered eligible for T-ICBT. An additional interest to this study, participants were asked to answer the yes or no question: “Have you had any problems with sleep, appetite/weight, energy level, or sexual activity?” following the symptom measures. If they responded “yes,” participants were presented with an open-ended comment box that asked them to provide details about the difficulties they had experienced in those areas.

After completion of the web-based screening questionnaire, participants were briefly interviewed by a clinician via the telephone to further assess suitability for the Wellbeing Course. Specifically, telephone screeners would provide more information about the nature of the Wellbeing Course to ensure clients were interested and ask any additional questions to determine whether patients should be excluded because they (1) were considered high risk for suicide or reported very severe symptoms (n=55), (2) were seeking primary treatment for another disorder (ie, obsessive compulsive disorder, posttraumatic stress disorder, bipolar disorder, psychotic symptoms, and alcohol or drug problems; n=21), (3) were receiving regular concurrent in-person psychotherapy (eg, receiving face-to-face cognitive behavioral therapy [CBT] in an outpatient setting) that is expected to address the client’s concern (n=6), or (4) did not meet the inclusion criteria (eg, age or lack of computer access or skills) described earlier (n=21).

All patients accepted for treatment received the same T-ICBT intervention, a program called the Wellbeing Course, which was developed by the eCentreClinic in Sydney, Australia. The Wellbeing Course has previously been studied in numerous clinical trials and found to be effective in reducing symptoms of anxiety and depression [4,6,8,9,26], but specific examination of how T-ICBT impacts sleep, sexual functioning, energy, and appetite/weight has not been undertaken. The Wellbeing Course is 8 weeks in duration and comprises 5 core lessons designed to target symptoms of depression and/or anxiety: (1) the cognitive behavioral model, (2) thought monitoring and challenging, (3) dearousal strategies and pleasant activity scheduling, (4) gradual exposure, and (5) relapse prevention. Each lesson consists of psychoeducation, instruction on cognitive behavioral strategies used for symptom reduction, case examples, and homework exercises. Participants also have access to additional resources on topics that are relevant to depression and anxiety (ie, assertiveness, communication, problem solving, worry time, and sleep). As they work through the course, patients receive therapist assistance in the form of weekly emails and occasional telephone calls. Patients complete the PHQ-9 and GAD-7 symptom measures again after completing the 8-week course.

Data Analysis

Qualitative content analysis was used to explore patients’ responses to the open-ended question about their problems with sleep, sexual functioning, energy, and appetite/weight [27]. A conventional or inductive approach was used, wherein the data were coded without the use of a preexisting coding guide [27]. The researchers involved with data analysis had all received didactic training in qualitative research methods, and two researchers had experience publishing qualitative research.

Data analysis was an iterative process that began with one undergraduate research assistant (AA) reading each response closely to obtain an initial impression of the data and engage in open coding, wherein basic codes representing each unit of meaning were derived (eg, difficulty falling asleep). Subsequently, the research assistant met with a PhD student in psychology (ME) to discuss initial impressions and create a preliminary coding guide of keywords and definitions. The research assistant then comprehensively coded all participant responses using the new coding guide. Next, several researchers (AA, ME, and HH) came together to sort the individual codes into meaningful themes and developed definitions for each theme (eg, the theme *difficulty falling asleep* was defined to capture any client statement that describes problems with initiating sleep). Finally, to ensure the accuracy of the coding, all participant responses were recoded by a second PhD student

in psychology (KG) using the finalized coding guide. The lead author resolved a small number of coding discrepancies between the first and second coders and completed a final review of the data. After coding was complete, the identified themes (Table 1) were analyzed using descriptive statistics to determine which concerns and symptoms were reported most frequently by participants.

In addition, to preliminarily explore the impact of T-ICBT on sleep, sexual functioning, energy, and appetite/weight from pre- to posttreatment, an item analysis was conducted on patients' responses on the GAD-7 and PHQ-9. Specifically, descriptive statistics were used to calculate the percentage of clients endorsing each item pretreatment as well as mean symptom change scores from pre- to posttreatment. Although we included all items in the analysis, we were particularly interested in the items related to sleep (PHQ-9 item 3: "trouble falling or staying asleep, or sleeping too much"), energy (PHQ-9 item 4: "feeling tired or having little energy"), and appetite (PHQ-9 item 5: "poor appetite or overeating"). Neither the GAD-7 nor the PHQ-9 included an item to assess weight or sexual functioning; thus, we were unable to determine if T-ICBT helped resolve these problems.

Table 1. Description of each theme along with subcategories, representative quotes, and frequencies (N=462).

Theme and description	Subcategories	Representative quotes	Frequency, n (%)
Sleep: concerns about quantity or quality of sleep	• Difficulty falling asleep • Difficulty staying asleep • Difficulty returning to sleep • Early waking • Difficulty waking up • Sleeping too much • Nightmares • General sleep problems	• "I have troubles falling asleep, staying asleep, and when I finally fall asleep it is time to wake up and I have trouble waking up." • "Most days [I] wake up feeling unrested, even if I got to bed at a decent time." • "[I] want to sleep all the time."	339 (73.4)
Energy or motivation: concerns about lack of energy, motivation, and self-care	• Decrease in energy • Feeling tired or exhausted • Mentally tired • Lacking motivation • Little or no exercise	• "I feel tired all the time. [...] I can't even muster energy to exercise or do things I used to like doing." • "I don't want to do anything except sleep or sit." • "I'm just very very tired, exhausted, low energy, and have no motivation."	320 (69.3)
Appetite/weight: concerns related to eating, appetite, and weight changes	• Variable appetite • Appetite greater than usual • Appetite less than usual • Poor diet • Emotional eating • Weight loss • Weight gain • Weight fluctuations • Overweight • Unsuccessful weight loss • Unsuccessful weight gain	• "Some days my appetite is very poor, some days I feel like eating all the time." • "I have no appetite and usually just eat because I have to." • "In the last year I have gained 25lbs due to depression and not being able to continue with my normal active lifestyle."	265 (57.4)
Sexual functioning: concerns with low sexual interest or engagement	• Low sex drive • Difficulty with arousal • Lack of sexual activity	• "My sex drive is pretty much non-existent." • "Sexually, [I am] unable to get aroused." • "I have been avoiding intimacy."	139 (30.1)
Cognitive-affective symptoms: concerns related to emotion regulation or cognitive functioning	• Anxiety • Depression • Angry and irritable • Low self-esteem • Emotional fluctuations • Lack of focus • Inability to relax • Cannot turn off mind • Constant worry • Overthinking	• "[I] cry regularly. [I am] irritable and moody." • "I lose concentration easily and feel foggy." • "[I] can't relax."	117 (25.3)
Somatic symptoms: concerns about physical symptoms and overall health	• Digestive problems • Nausea • Dizziness • General health concerns	• "I constantly feel sick [to my stomach]." • "[In] the evening I feel pressure behind my eyes and a slight headache. " • "[I am] always sick with colds and flus."	30 (6.5)

Results

Sample Characteristics

A total of 462 patients enrolled in the Wellbeing Course during the study period (see Table 2 for sample characteristics) and completed initial screening measures. Of the 442 patients who began treatment by logging in to the website, 341 (77.1%) completed 4 of the 5 core lessons and 298 (67.4%) completed all 5 lessons and completed posttreatment measures. Almost three-fourths of the participants were female (339/462, 73.4%), and the average age was 37 years (SD 12.33). The majority of participants reported being in a relationship and living in a city with more than 10,000 people. Participants varied widely in education level, although only 2.8% (13/462) had less than a high school education. Most participants reported being employed (283/462, 61.3%).

Table 2. Participant characteristics (N=462).

Variables	Values
Age (years)	
Mean (SD)	37.12 (12.33)
Range	18-86
Gender, n (%)	
Male	119 (25.8)
Female	339 (73.4)
Other	4 (0.9)
Marital status, n (%)	
Single or never married	104 (22.5)
Married	226 (48.9)
Living with partner	81 (17.5)
Separated or divorced	45 (9.7)
Widowed	6 (1.3)
Education, n (%)	
Less than high school	13 (2.8)
High school diploma	82 (17.7)
College certificate or diploma	135 (29.2)
Some university	64 (13.9)
University undergraduate degree	110 (23.8)
University professional or graduate degree	58 (12.5)
Employment status, n (%)	
Employed full time	224 (48.5)
Employed part time	59 (12.8)
Unemployed	38 (8.2)
Homemaker or retired or student	97 (21.0)
Short- or long-term disability	44 (9.5)
Location, n (%)	
Large city (>200,000)	185 (40.0)
Small to medium city (10,000-200,000)	153 (33.1)
Town or village	94 (20.3)
Farm	29 (6.3)
Reserve	1 (0.2)

Problems With Sleep, Sexual Functioning, Energy, and Appetite

The vast majority of patients (438/462, 94.8%) responded "yes" to the question asking if they had experienced problems related to sleep, sexual functioning, energy, or appetite/weight and thus responded to the open-ended comment box asking them to provide more detail about their problems in these areas. Patients described problems with sleep, sexual functioning, energy or motivation, and appetite/weight; however, they also reported on cognitive-affective symptoms and somatic symptoms. Table 1 includes descriptions, examples, representative quotes, and frequencies for each identified category. The most commonly reported problems were sleep, energy/motivation, and appetite/weight. Of the patients who responded to the open-ended comment box, 13.7% (60/438) provided a response that was coded into only 1 category, 24.8% (109/438) of responses were coded into 2 categories, 37.7% (165/438) were coded into 3 categories, 18.3% (80/438) were coded into 4 categories, 4.8% (21/438) were coded into 5 categories, and 0.9% (4/438) were coded into all 6 categories.

Treatment Outcomes and Item Analysis

Overall, patients' anxiety and depression scores decreased over the course of treatment. The mean total scores at the time of enrollment were 12.6 (SD 5.0) on the GAD-7 and 12.5 (SD 6.1) on the PHQ-9, which represents moderate symptoms of anxiety and depression. In this sample, treatment resulted in a mean score reduction at posttreatment of 6.6 points (SD 5.1) on the GAD-7 and 5.9 points (SD 5.1) on the PHQ-9, indicating that patients' anxiety and depression symptoms were in the mild range at the end of treatment. The results from the GAD-7 and PHQ-9 item analyses are presented in Table 3. Of the symptoms of interest in this study, low energy was the most frequently endorsed item on the PHQ-9, with 94.1% (428/455) of patients indicating they had felt tired or had little energy several days or more over the last 2 weeks. Sleep disruption was the second most commonly endorsed symptom, with 87.9% (400/455) of patients indicating they experienced trouble falling or staying asleep or sleeping too much on several days or more over the last 2 weeks. Appetite problems were another commonly endorsed item on the PHQ-9, with 79.3% (361/455) of patients indicating they experienced poor appetite or overeating several days or more over the last 2 weeks. Promisingly, T-ICBT appeared to have a positive impact on sleep, energy, and appetite symptoms, that is, patients reported a decrease in the frequency of low energy, sleep disturbance, and appetite problems from pre- to posttreatment. It is important to note, however, that patients still reported experiencing symptoms in these areas posttreatment. The lack of GAD-7 or PHQ-9 items related to sexual functioning and weight meant it was not possible to assess whether T-ICBT led to improved functioning in those areas.

Table 3. Frequencies of responses to the 7-item Generalized Anxiety Disorder scale and the 9-item Patient Health Questionnaire items at pretreatment (N=462).

Item and symptom	Distribution of pretreatment responses				Pretreatment mean score (SD)	Posttreatment mean score (SD)
	0 (not at all), n (%)	1 (several days), n (%)	2 (more than half the days), n (%)	3 (nearly every day), n (%)		
7-item Generalized Anxiety Disorder (n=454)						
Feeling nervous or anxious	12 (2.6)	127 (28.0)	143 (31.5)	172 (37.9)	2.05 (0.87)	0.99 (0.80)
Uncontrollable worrying	27 (5.9)	125 (27.5)	129 (28.4)	173 (38.1)	1.99 (0.95)	0.83 (0.88)
Worrying about a variety of things	13 (2.9)	114 (25.1)	137 (30.2)	190 (41.9)	2.11 (0.88)	0.95 (0.89)
Trouble relaxing	40 (8.8)	130 (28.6)	130 (28.6)	154 (33.9)	1.88 (0.98)	0.81 (0.90)
Restlessness	145 (31.9)	162 (35.7)	91 (20.0)	56 (12.3)	1.13 (1.00)	0.52 (0.80)
Irritability	24 (5.3)	142 (31.3)	128 (28.2)	160 (35.2)	1.93 (0.94)	1.02 (0.97)
Fearing something awful	96 (21.1)	144 (31.7)	109 (24.0)	105 (23.1)	1.49 (1.07)	0.60 (0.86)
9-item Patient Health Questionnaire (n=455)						
Anhedonia	54 (11.9)	201 (44.2)	107 (23.5)	93 (20.4)	1.53 (0.95)	0.76 (0.80)
Feeling depressed and hopeless	51 (11.2)	203 (44.6)	117 (25.7)	84 (18.5)	1.51 (0.92)	0.76 (0.79)
Sleep disruption	55 (12.1)	122 (26.8)	118 (25.9)	160 (35.2)	1.84 (1.04)	1.00 (0.93)
Low energy	27 (5.9)	127 (27.9)	128 (28.1)	173 (38.0)	1.98 (0.95)	1.12 (0.90)
Poor appetite or overeating	94 (20.7)	133 (29.2)	111 (24.4)	117 (25.7)	1.55 (1.09)	0.81 (0.93)
Feeling bad about self	59 (13.0)	163 (35.8)	122 (26.8)	111 (24.4)	1.63 (0.99)	0.69 (0.85)
Trouble concentrating	100 (22.0)	147 (32.3)	117 (25.7)	91 (20.0)	1.44 (1.04)	0.62 (0.83)
Psychomotor agitation or retardation	236 (51.9)	124 (27.3)	67 (14.7)	28 (6.2)	0.75 (0.92)	0.31 (0.60)
Self-harm or suicidality	359 (78.9)	69 (15.2)	19 (4.2)	8 (1.8)	0.29 (0.63)	0.11 (0.38)

Discussion

Principal Findings

This formative study qualitatively examined patients' self-reported concerns with sleep, sexual functioning, energy, and appetite/weight at the time of enrollment in T-ICBT for depression and anxiety to learn more about the prevalence and nature of patients' problems in these areas. In addition, an item analysis was conducted on screening measures of anxiety and depression to explore whether T-ICBT produces improvements in these symptoms posttreatment. Of importance, the vast majority of T-ICBT clients (438/462, 94.8%) endorsed problems in at least one area and 87.4% (404/462) of clients reported problems in more than one area. Although this is not surprising given research has demonstrated that sleep problems, sexual dysfunction, lack of energy, and appetite/weight changes are shared features of anxiety and depression [13-20], the extent of the problems is important to understand so as to inform T-ICBT delivery in routine care.

Qualitative analysis of patients' descriptions of their concerns revealed that sleep disturbance was the most common problem described by clients (339/462, 73.4%). As hypothesized, clients' sleep concerns were consistent with insomnia (eg, difficulty falling asleep, difficulty staying asleep, and early waking) and hypersomnia (eg, difficulty waking up, and sleeping too much), although a number of clients also reported having nightmares. Loss of energy or motivation was the second most common client concern (320/462, 69.3%). Consistent with our expectations, clients described experiencing a loss of energy and mental or physical tiredness; however, they also reported a lack of motivation and loss of interest in self-care. Interestingly, clients often discussed energy and motivation in tandem, which is why these problems are grouped together in the results. The third most common client concern was changes in appetite/weight (265/462, 57.4%), with some clients describing an increase in appetite/weight and others a decrease in appetite/weight. This is mostly consistent with our hypothesis, but we did not fully anticipate how widespread problems would be with poor diet or emotional eating. A less common concern, although still affecting one-third of the sample, was sexual dysfunction (139/462, 30.1%). In this domain, clients reported loss of sexual desire and difficulty with arousal, which is in line with our hypothesis.

Although the question presented to clients asked specifically about sleep, sexual functioning, energy, and appetite/weight, clients also described problems in other areas without being prompted to do so. Specifically, 25.3% (117/462) of clients reported experiencing cognitive-affective symptoms (eg, worry,

low mood, anger or irritability, lack of focus, and inability to relax), and 6.5% (30/462) of clients described somatic symptoms (eg, digestive problems, nausea, dizziness, and compromised immune system). These concerns were not the intended focus of this study, yet the fact that they were brought up by clients suggests they are important. It remains to be seen if these concerns would be more prevalent if the question prompted clients to think about these domains.

The results of the item analysis provided preliminary evidence that T-ICBT leads to improvements in sleep, energy, and appetite in 8 weeks. This promising finding suggests that T-ICBT has a positive impact on problems in these areas. It is important to note, however, that patients' problems with sleep, energy, and appetite were not eliminated entirely, which suggests that it may still be possible to optimize T-ICBT to better address these problems. Moreover, we were unable to monitor changes in sexual functioning or weight over the course of treatment because these symptoms were not represented by the items on the GAD-7 of PHQ-9. This seems especially problematic for sexual functioning, an issue that was raised by more than one-third of the study sample. A recent study found that face-to-face CBT resulted in modest reductions in reports of sexual dysfunction [20], but it is unknown if T-ICBT produces similar reductions in sexual concerns.

The findings of this study have implications for future T-ICBT research with regard to symptom screening and outcome monitoring. Specifically, single-item measures can provide some indication of symptom change over the course of treatment, but it is unclear whether they are sensitive enough to the wide variety of problems patients reported with regard to sleep, energy, appetite/weight, and sexual functioning. Thus, there may be value in expanding screening or outcome measures to include more through assessment of sleep, sexual functioning, energy or motivation, and appetite/weight. Enhancing existing screening or outcome measures would allow researchers to make concrete conclusions about the efficacy of T-ICBT in improving these symptoms. However, it would be important to balance the desire for more detailed information with the increased burden on clients when completing screening and posttreatment questionnaires. If measures were to be added, they would need to be brief, psychometrically sound, and not redundant with existing items or measures. In the long term, learning about the effects of T-ICBT on sleep, sexual functioning, energy or motivation, and appetite/weight might provide insight into ways to enhance treatment content, which, in turn, might influence treatment engagement or outcomes. The results of this study also have implications for clinical practice and future T-ICBT program development. With regard to therapist support, the results of this study emphasize that therapists need to be aware of how common and diverse problems with sleep, energy or motivation, appetite/weight, and sexual functioning are within their client group and provide support accordingly. With respect to program development, the results of this study suggest that concerns related to sleep, energy or motivation, appetite/weight, and sexual functioning are common, and ensuring program content relates to these concerns (eg, by using symptoms of sleep disruption in case examples provided to patients) could be an important part of maximizing the patient acceptability of T-ICBT.

Strengths and Limitations

This study makes a valuable contribution to the literature because it is the first to qualitatively examine the prevalence and nature of T-ICBT clients' self-described problems with sleep, sexual functioning, energy, and appetite/weight. The qualitative approach to data analysis provided a depth of information unavailable in previous quantitative studies. For example, we found that clients' specific concerns about sleep, sexual functioning, energy or motivation, and appetite/weight are diverse and may not be adequately captured by common symptom screening measures. The inclusion of a large sample of clients increases our confidence that the results are generalizable to other clients seeking T-ICBT in routine care. The addition of the item analysis is another notable strength of the study, as it provides preliminary evidence that T-ICBT has some positive effects on sleep disruption, low energy, and appetite change.

This study also had several limitations. First, the wording of the question patients responded to may have prompted clients to respond in a certain way. Patients were specifically asked only a single question about sleep, sexual functioning, energy, and appetite/weight, and using a more open-ended question or asking about each symptom domain separately may have produced different results. Second, in terms of question order, clients responded to the research questions after they had completed the GAD-7 and PHQ-9, which makes it possible that clients were primed by the items related to sleep, energy, appetite, and cognitive-affective symptoms. In interpreting the results, readers should also consider that the PHQ-9 and GAD-7 measure the frequency of symptoms but not the severity of the symptoms, and there is no way to know from our data what proportion of clients were experiencing clinically significant difficulties in each area. It is also important to note that the characteristics of the sample used in this study may limit the generalizability of the results to other ICBT programs. The characteristics of the sample that should be considered by the reader when interpreting the results include the fact that a large proportion of patients were self-referred to the service (193/462, 41.8%), that 58.9% (272/462) of patients reported receiving simultaneous pharmacological treatment for their concerns at pretreatment, and that patients who were receiving simultaneous in-person therapy (n=6) were excluded from the study. Posttreatment results must also be interpreted with caution, as 32.6% (144/442) of participants had withdrawn from treatment before the 8-week point, and the extent to which patients used skills described in the course or gained knowledge was not assessed in this study.

Conclusions

This study presents an overview of the concerns related to sleep, energy, appetite, and sexual functioning reported by individuals seeking T-ICBT for anxiety and depression and may therefore be of interest to program designers and clinicians interested in offering T-ICBT. We found that the majority of clients reported symptoms such as sleep disruption, lack of energy or motivation, and appetite/weight changes, and just less than one-third of the

sample reported experiencing sexual dysfunction. We also found preliminary evidence that T-ICBT resulted in reductions in the reported frequency of sleep disruption, low energy, and appetite disturbance over 8 weeks. The results of this study allowed us to identify several areas where T-ICBT outcome measures could be improved and provided a number of directions for future research.

Acknowledgments

This work was supported by funding provided by the Canadian Institutes of Health Research (reference numbers 293379 and 152917), Saskatchewan Health Research Foundation, and Rx & D Health Research Foundation. The authors would also like to thank Luke Schneider for sharing his expertise in qualitative research methodology during the early phases of this project.

Conflicts of Interest

None declared.

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. Geneva: World Health Organization; 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2020-09-29]
2. Păsărelu CR, Andersson G, Nordgren LB, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
3. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
4. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
5. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. J Am Med Assoc Psychiatry 2017 Sep 1;74(9):875-884 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
6. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
7. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Jul 26;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
8. Fogliati VJ, Dear BF, Staples LG, Terides MD, Sheehan J, Johnston L, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided internet-delivered treatment for panic disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2016 Apr;39:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2016.03.005] [Medline: 27003376]
9. Dear BF, Staples LG, Terides MD, Fogliati VJ, Sheehan J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for social anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2016 Aug;42:30-44 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.004] [Medline: 27261562]
10. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
11. Newby JM, Mewton L, Andrews G. Transdiagnostic versus disorder-specific internet-delivered cognitive behaviour therapy for anxiety and depression in primary care. J Anxiety Disord 2017 Mar;46:25-34. [doi: 10.1016/j.janxdis.2016.06.002] [Medline: 27396841]
12. Titov N, Hadjistavropoulos HD, Nielssen O, Mohr DC, Andersson G, Dear BF. From research to practice: ten lessons in delivering digital mental health services. J Clin Med 2019 Aug 17;8(8):1239 [FREE Full text] [doi: 10.3390/jcm8081239] [Medline: 31426460]
13. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Arlington, VA: American Psychiatric Publishing, Inc; 2013.

14. Taylor DJ, Lichstein KL, Durrence HH, Reidel BW, Bush AJ. Epidemiology of insomnia, depression, and anxiety. Sleep 2005 Nov;28(11):1457-1464. [doi: 10.1093/sleep/28.11.1457] [Medline: 16335332]
15. Rosekind MR, Gregory KB. Insomnia risks and costs: health, safety, and quality of life. Am J Manag Care 2010 Aug;16(8):617-626 [FREE Full text] [Medline: 20712395]
16. Demyttenaere K, De Fruyt J, Stahl SM. The many faces of fatigue in major depressive disorder. Int J Neuropsychopharmacol 2005 Mar;8(1):93-105 [FREE Full text] [doi: 10.1017/S1461145704004729] [Medline: 15482632]
17. Privitera GJ, Misenheimer ML, Doraiswamy PM. From weight loss to weight gain: appetite changes in major depressive disorder as a mirror into brain-environment interactions. Front Psychol 2013 Nov 21;4:873 [FREE Full text] [doi: 10.3389/fpsyg.2013.00873] [Medline: 24312070]
18. Zemishlany Z, Weizman A. The impact of mental illness on sexual dysfunction. Adv Psychosom Med 2008;29:89-106. [doi: 10.1159/000126626] [Medline: 18391559]
19. Tsuno N, Besset A, Ritchie K. Sleep and depression. J Clin Psychiatry 2005 Oct;66(10):1254-1269. [doi: 10.4088/jcp.v66n1008] [Medline: 16259539]
20. Hoyer J, Uhmann S, Rambow J, Jacobi F. Reduction of sexual dysfunction: by-product of cognitive-behavioural therapy for psychological disorders? Sex Relation Ther 2009 Feb;24(1):64-73. [doi: 10.1080/14681990802649938]
21. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. Br Med J 1995 Jul 1;311(6996):42-45 [FREE Full text] [doi: 10.1136/bmj.311.6996.42] [Medline: 7613329]
22. Watts S, Newby JM, Mewton L, Andrews G. A clinical audit of changes in suicide ideas with internet treatment for depression. Br Med J Open 2012 Sep 13;2(5):e001558. [doi: 10.1136/bmjopen-2012-001558] [Medline: 22983787]
23. Hadjistavropoulos HD, Schneider LH, Mehta S, Karin E, Dear BF, Titov N. Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. J Anxiety Disord 2019 Apr;63:51-60 [FREE Full text] [doi: 10.1016/j.janxdis.2019.02.002] [Medline: 30844613]
24. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
25. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
26. Hadjistavropoulos HD, Nugent MM, Alberts NM, Staples L, Dear BF, Titov N. Transdiagnostic internet-delivered cognitive behaviour therapy in Canada: an open trial comparing results of a specialized online clinic and nonspecialized community clinics. J Anxiety Disord 2016 Aug;42:19-29. [doi: 10.1016/j.janxdis.2016.05.006] [Medline: 27244278]
27. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

Abbreviations

CBT: cognitive behavioral therapy
GAD-7: 7-item Generalized Anxiety Disorder
ICBT: internet-delivered cognitive behavioral therapy
PHQ-9: 9-item Patient Health Questionnaire
T-ICBT: transdiagnostic internet-delivered cognitive behavioral therapy

Edited by G Eysenbach, J Li; submitted 17.06.19; peer-reviewed by E Lee, BT Tulbure, C Păsărelu; comments to author 08.04.20; revised version received 06.08.20; accepted 07.09.20; published 13.10.20.

Please cite as:
Edmonds MR, Hadjistavropoulos HD, Gullickson KM, Asmundson AJN, Dear BF, Titov N
Understanding Problems With Sleep, Sexual Functioning, Energy, and Appetite Among Patients Who Access Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy for Anxiety and Depression: Qualitative Exploratory Study
JMIR Form Res 2020;4(10):e15037
URL: http://formative.jmir.org/2020/10/e15037/
doi:10.2196/15037
PMID:33048054

Original Paper

XML Data and Knowledge-Encoding Structure for a Web-Based and Mobile Antenatal Clinical Decision Support System: Development Study

Ever Augusto Torres Silva[1*], BME, MSc; Sebastian Uribe[1*], BME, MSc; Jack Smith[2*], PhD; Ivan Felipe Luna Gomez[1*], BME, MSc; Jose Fernando Florez-Arango[3*], PhD

[1]Bioengineering Research Group, Universidad Pontificia Bolivariana, Medellin, Colombia

[2]Department of Microbial Pathogenesis and Immunology, Texas A&M Unversity, College Station, TX, United States

[3]Department of Humanities in Medicine, Texas A&M University, Bryan, TX, United States

[*]all authors contributed equally

Corresponding Author:
Ever Augusto Torres Silva, BME, MSc
Bioengineering Research Group
Universidad Pontificia Bolivariana
Calle 78B #72A - 109
Medellin, 050034
Colombia
Phone: 57 44488388 ext 19337
Email: ever.torres@upb.edu.co

Abstract

Background: Displeasure with the functionality of clinical decision support systems (CDSSs) is considered the primary challenge in CDSS development. A major difficulty in CDSS design is matching the functionality to the desired and actual clinical workflow. Computer-interpretable guidelines (CIGs) are used to formalize medical knowledge in clinical practice guidelines (CPGs) in a computable language. However, existing CIG frameworks require a specific interpreter for each CIG language, hindering the ease of implementation and interoperability.

Objective: This paper aims to describe a different approach to the representation of clinical knowledge and data. We intended to change the clinician's perception of a CDSS with sufficient expressivity of the representation while maintaining a small communication and software footprint for both a web application and a mobile app. This approach was originally intended to create a readable and minimal syntax for a web CDSS and future mobile app for antenatal care guidelines with improved human-computer interaction and enhanced usability by aligning the system behavior with clinical workflow.

Methods: We designed and implemented an architecture design for our CDSS, which uses the model-view-controller (MVC) architecture and a knowledge engine in the MVC architecture based on XML. The knowledge engine design also integrated the requirement of matching clinical care workflow that was desired in the CDSS. For this component of the design task, we used a work ontology analysis of the CPGs for antenatal care in our particular target clinical settings.

Results: In comparison to other common CIGs used for CDSSs, our XML approach can be used to take advantage of the flexible format of XML to facilitate the electronic sharing of structured data. More importantly, we can take advantage of its flexibility to standardize CIG structure design in a low-level specification language that is ubiquitous, universal, computationally efficient, integrable with web technologies, and human readable.

Conclusions: Our knowledge representation framework incorporates fundamental elements of other CIGs used in CDSSs in medicine and proved adequate to encode a number of antenatal health care CPGs and their associated clinical workflows. The framework appears general enough to be used with other CPGs in medicine. XML proved to be a language expressive enough to describe planning problems in a computable form and restrictive and expressive enough to implement in a clinical system. It can also be effective for mobile apps, where intermittent communication requires a small footprint and an autonomous app. This approach can be used to incorporate overlapping capabilities of more specialized CIGs in medicine.

(JMIR Form Res 2020;4(10):e17512) doi:10.2196/17512

KEYWORDS

clinical decision support systems; computer-interpretable guidelines; knowledge representation; state machine; system design; XML

Introduction

Background and Significance

With the increasing adoption of electronic health records and hospital information systems, the implementation of clinical practice guidelines (CPGs) through integration with these information systems is possible during clinical encounters [1], creating what is known as clinical decision support systems (CDSSs), which assist physicians during health care encounters. CDSSs attempt to mimic the way humans use clinical guidelines with patient information and make decisions based on existing clinical knowledge and knowledge specific to a patient [2]. These systems are built using representations of knowledge and information about different diseases, treatment protocols, findings, and interpretations [3].

A CDSS produces patient-specific output based on patient data combined with these representations of medical knowledge [4]. Traditionally, three functions are supported by a CDSS: the provision of automated clinical information management such as data entry and retrieval, the attention-focusing functions such as medical alerts and reminders, and the provision of patient-specific recommendations or advice based on individual patient data [5]. All these functions can be useful in providing health care services, but the execution and effectiveness of these functions are determined by the effectiveness of the utilization of medical knowledge in the associated clinical workflow [3].

There have been widespread reports of dissatisfaction and user complaints with existing CDSS functionality, and satisfaction is considered one of the primary challenges for improving CDSS development and acceptance [6]. A special challenge in this regard is implementing a CDSS system with interactions that match actual and desired clinical workflow as closely as possible. This requires that the CDSS interactions with the user be context sensitive and accessible at the point of care [7]. More importantly, studies have shown that a CDSS should also integrate into the overall organizational workflow in order to make its use easy and efficient [4].

In a CDSS, computer-interpretable guidelines (CIGs) are used to formalize medical knowledge contained in CPGs into a computable form. Additionally, we can create a set of software functions for the CDSS user interactions that also match the clinical workflow [6]. However, existing CIG frameworks require a specific interpreter or compiler for each computable representation language, hindering their widespread implementation and interoperability [7]. This has led us to propose a different approach to CIGs and workflow representation that addresses how both clinical knowledge and data are represented while retaining the ability to capture workflow constraints in order to create a more positive user perception of the CDSS. We also desired to maintain a small software and communications footprint for web applications and mobile apps. Our approach was originally intended to create a readable and minimal syntax for a web and mobile application CDSS for antenatal care guidelines that would improve human-computer interaction and enhance usability by aligning the system behavior with clinical workflow. In the antenatal system design, a requirement was to implement a CDSS that was well integrated with existing medical workflow, first as a web application and then as a mobile app. A number of such systems have been developed and shown to improve clinical outcomes [8-10].

Alternative Solutions

We considered several different representation languages for our CDSS system. One approach we considered was to use an existing CIG syntax, such as Arden syntax or Guideline Interchange Format, to represent the medical knowledge found in the CPG. However, we decided to develop our own syntax to handle clinical data and knowledge because we are aiming to implement our CDSS with a model-view-controller (MVC) architecture appropriate for both web applications and mobile apps. We intended to repurpose the CIG knowledge engine core and the represented medical content for mobile app development. We had the additional goal of ensuring that the executable representation be human readable and have sufficient but minimal syntax and process elements for our particular CPGs. Importantly, we took a minimal design approach, in that we wanted the footprint of the software to have as small a computational and data footprint as possible, since we intended to use the resulting tools both in a web application and mobile app. In our environment, the mobile app may not have continuous connectivity to the networked computing resources, requiring it to execute autonomously if necessary.

With these requirements in mind, we chose XML as a better language to represent CPG knowledge. For web applications and mobile apps, this is an ideal choice as a core component of the World Wide Web. XML is key to formatting content into HTML pages and is an industry standard for data communication among different computer systems [11]. We took advantage of the XML schema over other formats like JavaScript Object Notation to validate the computable representation of CPG documents and verify each piece of item content in a document [12].

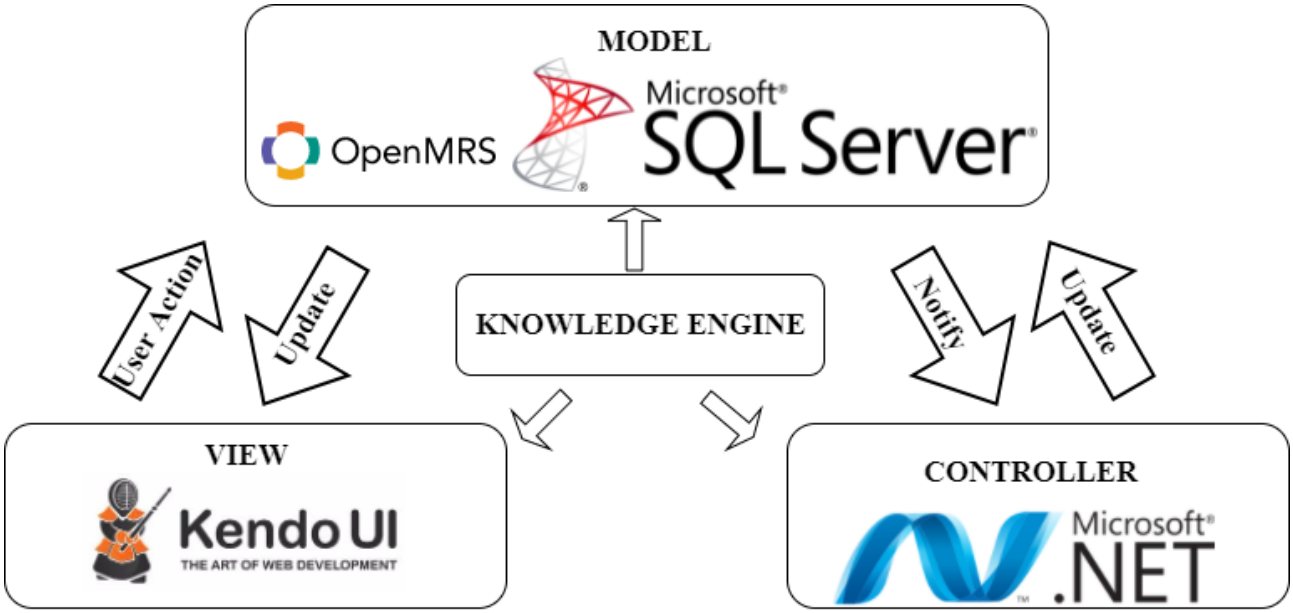
There is increasing interest in software frameworks and the feasibility of representing knowledge such as CPGs in semantic web technologies [13-18]. One such advanced system combines a formalization of CPGs using fuzzy cognitive maps (FCMs) implemented in semantic web technology [10]. The CPGs are represented in the form of if-then fuzzy rules. The representation of FCMs uses Notation 3 [11], which is a shorthand non-XML serialization of resource description framework models. The Euler sharp reasoning engine [12] is used to implement inferencing for this FCM implementation. This approach, as well as similar ones that use a sematic web technology layer, has great expressive power but is in conflict with our minimalist requirements for mobile environments, in which connectivity

is intermittent and thus requires that CDSS run with minimal computing resources without connectivity to the internet.

There are many other challenges to CDSSs in the environments in which we are implementing our system. There are issues of scale in the management of guidelines in CIG forms, history tracking, CIG version control, and automatic aggregation of CIGs. In this paper, we focus on a narrow subset of such desirable requirements. We are only concerned with having sufficient expressivity of the representation while maintaining a small communication and software footprint for mobile apps that can also be repurposed without modification for incorporation into web applications and other mobile apps. To our knowledge, none of the approaches taken above would be satisfactory for those requirements. In the future, as our approach develops, we will also address these other important issues and be in a better place to compare our approach on those dimensions. In this paper, we focus on the adequacy of satisfying the requirements for expressivity and minimal communication and computing resources.

In Figure 1, we illustrate the architecture design of our CDSS, which uses the MVC architecture and a knowledge engine based on XML. Specifically, we have (1) a model, (2) a view, (3) a controller, and (4) a knowledge engine.

Figure 1. Architecture design used in the web-based clinical decision support system.



First, the model consists mainly of a database system used to store the patient's phenotypic and laboratory data so the database can be queried when required [19]. In our CDSS web application, we use Microsoft SQL Server (Microsoft Corp) as the relational database management system with a data model based on the OpenMRS (OpenMRS Inc) medical record system.

Second, the view refers to the system components the user interacts with that display in the user interface.

Third, the controller refers to the system components responsible for processing queries and the user's interactions. The controllers pass patient input data to the computable guideline model, which is used to query the database, and then select a view to display for the user interface [19].

Fourth, the knowledge engine is the CIG where the decision-making rules capturing the CPGs are stored and where the knowledge for integrating the clinical health care workflow and CPGs are stored and interpreted.

In this paper, we exclusively present the design of the knowledge engine, a collection of XML documents and an associated interpreter to handle clinical data and encoded knowledge. The relation between the functionality of the CIG subsystem and the MVC architecture involves the knowledge engine driving the controller's functionality, alongside a graphic representation of the views. We analyzed some CIG syntaxes in order to find common and minimal features for knowledge representation to incorporate in our XML. We particularly wanted to have design control of features directing user attention to system outputs and improve system usability by being able to match the standard workflow of the clinical situation.

Methods

Knowledge Engine Background

For the CDSS knowledge engine, we adopted state machines, artificial intelligence (AI) planning techniques and methods, and minimal CIG functionality. We now briefly describe our decision-making process for each of these.

State Machine Modeling of Behavior for Health Care

There are many ways to computationally model knowledge and behavior in AI systems, and the use of state machines is one of the oldest and best known [14]. State machines model a system's states, or features of a system, at a particular point in time and characterize its future behavior based on these states [15]. In our approach, we use state machine models as an intermediate level of design and implementation before representing the state machines in executable XML code.

Planning

In AI, a plan is generally defined as a sequence of actions that will achieve a specific state. A state refers to the multiple logical conditions that are true in a certain situation, or what could be also known as the "state of the world" [16]. Planning can be used to take advantage of the knowledge of the world. Knowledge in AI refers to information and conditions about the world and how actions change and affect the world [17]. AI planning approaches can be used to represent what is known about the current state of the world and the available actions. In terms of our CDSS, the world would be the patient's situation, and the actions would be the testing, observations, diagnoses, and treatments that change the state of the patient. An AI planning conceptualization of CIG allows for a framework that incorporates the various approaches to CIG taken in Table 1 and our approach. Adopting this approach allows us to envision using AI planning techniques and methods for this part of the CDSS.

Table 1. Computer-interpretable guideline formalisms comparison.

Detail	Arden syntax	GLIF[a]	PROforma	Asbru	Eon
Model	Medical logic module	Object-oriented flowchart structured in steps	PROforma task ontology	Plan	Dharma guideline model
Model elements	Maintenance, knowledge, library, resources	Branch, decision, action, patient state steps	Plans, decisions, actions, inquiries	Preferences, intentions, conditions, effects	Scenarios, decisions, actions, activities
Language	- MLMs[b] are text based (each MLM is encoded as an ASCII[c] file)	- UML[d] class diagrams in GLIF3 XML-based syntax - RDF[e] language	- Guidelines are translated into language called LR2L - Contains a formal expression language	- DTD[f] in Backus-Naur form - Control-flow language are defined by means of XML	RDF

[a]GLIF: Guideline Interchange Format.

[b]MLM: medical logic module.

[c]ASCII: American Standard Code for Information Interchange.

[d]UML: Unified Modeling Language.

[e]RDF: Resource Description Framework.

[f]DTD: document type definition.

Minimal CIG Functionality Objective

In designing a minimal-specification XML file to manage clinical data and knowledge for a web-based and mobile CDSS for antenatal care, we compared and analyzed some CIG approaches found in OpenClinical [20] (see "CIG Comparisons"). We will not attempt to explain each of the methods here, as those details can be found in the original papers [21-25]. We incorporated into our representation the elemental model structures and language of each CIG.

The knowledge engine design also integrated the requirement of matching care workflow that was desired in the CDSS. Here we used a work ontology analysis [26] of the CPG, as it is applied to antenatal care in our particular target clinical settings. The ontology analysis was used to reach a common understanding of the structure of information [27] (see "Knowledge Extraction") in the care process, since the format of CPGs is not standardized and shows variations according to the organization producing the guidelines and the clinical area [28]. This knowledge extraction and ontology analysis were performed for antenatal health care from published clinical practice guidelines related to pregnancy and childbirth in various sources [1,29-32]. The knowledge and ontology analysis found in the CPGs were then aligned with AI planning theory conceptualizations of the clinical workflow in antenatal care in order to establish the knowledge representations in the knowledge engine.

We then used Microsoft Visual Studio (Microsoft Corp) to develop an XML Schema Definition (XSD) to ease the knowledge-encoding process into rule statements. Lastly, we compared our XML document file with an Arden syntax file that we coded doing a similar task in the interest of comparing our representations with one of the CDSS standards often used to manage and apply clinical knowledge in health care settings.

Results

CIGs Comparison

Table 1 displays the key elements of some guideline formalism models. All the approaches support a basic AI planning structure to handle decisions and actions based on medical criteria. At a conceptual level, they are very similar, but the formalisms share similarities and differences in form and terminology. They all abstractly incorporate the AI conceptualization of planning as being capturable by a hierarchical task structure, that is, steps in a plan that can be represented by representations of their preconditions, actions, and goals. The goal itself refers to the values of states of the world or knowledge states that must be attained to satisfy a medical decision criterion that would lead to a certain action. The preconditions specify the context in which an encapsulated collection of represented knowledge (a knowledge module) should be executed, and the action to be taken is typically an action to assist the user (physician or health care worker).

Knowledge Extraction

As discussed in de Clercq et al [22], a balance must be maintained between the aspects of abstractness, expressiveness, formalization, acquisition, and execution of the knowledge in order to create a successful CIG. Real effectiveness is also dependent on the guideline development and knowledge extraction processes from original sources. We implemented the framework presented in Boxwala et al [33] to extract the knowledge found in the guidelines. We separated the main health care actions into the main actions that health care professionals could perform. Table 2 illustrates some of the results of our ontology and knowledge analysis of the antenatal CPG knowledge. This table is structured into the divisions of knowledge represented in a basic AI planning system of the applicable medical procedures, treatments, diseases, and relevant information in our CPGs. It is important to note that our knowledge representation framework, techniques, and methods were designed in a way that can be a used for other medical system design purposes. The current implementation is a proof of concept that the general framework can be successfully specialized for a specific set of medical problems to be represented in a CDSS.

Table 2. Ontology knowledge structure extraction for antenatal clinical practice guidelines.

Main actions	Health care provider action	Goal	Precondition	Action
Observation	Examination findings, family history, lifestyle factors, health summary, pregnancy summary, nutrition summary, physical activity summary, social summary, laboratory test, reason for encounter	Record observation and information about the patient	Information is missing or incomplete	Request information
Evaluation	Absence of information, clinical synopsis, adverse reaction risk, health risk assessment, problem/diagnosis	Analyze information and absences of information	Information is in database	Compare information, risk assessment
Action	Care plan, health education, medication order, laboratory test request, procedure request, notification	Suggest a proper care plan and notify health care provider of an event	Evaluation is done	Suggest medication order, laboratory test, or procedure; generate a notification, alert, or message

XSD Design

At the next level of system design, we used XML-encoded rules to represent the state transitions of a state machine. This state machine protocol file is the core component for future proposed implementations of CDSS content captured in a state machine and interpreted as state machine behavior.

The CDSS uses XML logical methods to interpret data entries through conditional rules to achieve representations of goals and actions. The XML interpreter will usually execute the logic through a match-and-resolve process. First, all rules with conditions corresponding with the input data are evaluated, and if the conditions are satisfied, the rules will execute.

In the pseudocode of Figure 2 for the XML layer, it can be seen how our interpreter will read the XML file. The structure is designed so that the interpreter will use a small number of recursive functions to read a large amount of code. This can be seen in the pseudocode, where each time an evaluation is found, it calls a function to evaluate the logical value of the conditions.

Figure 2. Pseudocode for our knowledge representation.

```
W:= {State,Protocol};

EffectList[]:=empty;

run(W)

{

foreach (Node in Protocol)

        If (!Node.Goal) then

          foreach (Action in Node)

            If (evaluatePrecontion(Action.Precondition)) then

              Foreach (Effect in Action)

                If (Effect !in EffectList[Effect.type]) then

                  EffectList[Effect.type][] := Effect

      Update S

}

bool evaluate Precondition (operator1, operator2, operand)

{

      operator1status = operator2status= null

      if (operator1.iscomplex())

        operator1status = evaluatePrecondition (operator1)

      if (operator2.iscomplex())

        operator2status = evaluatePrecondition (operator2)

return eval
```

The rule sets are arranged in a hierarchical structure to capture basic AI planning concepts; inner rules will only evaluate if outer rules in the hierarchy are true. Nodes are aggregates of conditions, which are sets of data as well as a set of rules. Each XML node also has a goal representation, a precondition representation, and an action representation. An XML node corresponds to a basic AI planner module. After the execution of the actions, there will be an update of the representation in terms of state representations, as seen in the pseudocode.

Nodes are the containers where a particular set of goals are established; this is used to give a structure and readability to the XML file.

Goals are considered the main representational purpose for the node to achieve, and when the representation conditions are accomplished, the node will be considered complete and will no longer be able to execute. This is so the system will not compute statements for an undefined period of time. For each of the conditions or rules inside the goal representations and precondition representations, we declare a Boolean expression with the logical operator or operators, including the set operation, in a set as well as in range. Inside each Boolean expression, we reference a represented semantic path to a symbol to access the required values of concept representations that can be related to a specific ID in the database.

Preconditions are the statements that are required to be satisfied before a rule triggers the execution of the associated rule action. Even if the rule inside the goal satisfies the conditions, the action will not be executed. In our knowledge representation, actions can be triggered in 4 different ways: (1) the passage of time;

(2) the entry of a data pattern for a patient's symptom, problem, or diagnosis; (3) the entry of a representation of a diagnostic test value; and (4) the entry of a representation of a treatment result.

Actions are the representation's output based on specific input conditions, which in turn are based on the rule by which goals and preconditions are satisfied. The actions that are applicable to a state are all those whose preconditions are satisfied. An action means there is a transition between states for the state machine, meaning that certain medical conditions were satisfied in the goal and precondition nodes. An action might be a new request for data, a message, a medical alert, a medical recommendation, or even a calculation, depending on the situation. At this point, it is important to make it clear that when an action takes place, the conditions in the goal will be completed. Otherwise, the node will execute every time the system evaluates the conditions.

The result of our knowledge representation is written and stored in an XML file. XML-encoded rules represent state transitions of the state machine system. This XML file and associated interpreter will be the inference engine core for any expanded future CDSS CIGs with state machine behavior in our future designs.

XSD Structure

An XSD schema file was designed with the intention to ease the creation of the XML representations of CPGs with the medical knowledge representation proposed. In Figure 3, the final structure for our knowledge representation is presented. The most important data types of the structure are (1) Boolean expressions, (2) symbols, (3) values, (4) InSets, and (5) Parens.

First, the Boolean expression can be any of the logical operators: "and", "or", or "not". It may also be any of the comparison operators: equal to, greater than, less than, not equal to, greater than or equal to, or less than or equal to. It can be true or false. It also includes the set operations "in-set" and "in-range."

Second, a symbol is just a string representation of a variable.

Third, a value is the different "values" that symbols can take.

Fourth, the InSet type represents whether a particular symbol is in the set of values. The set can either be empty or infinite.

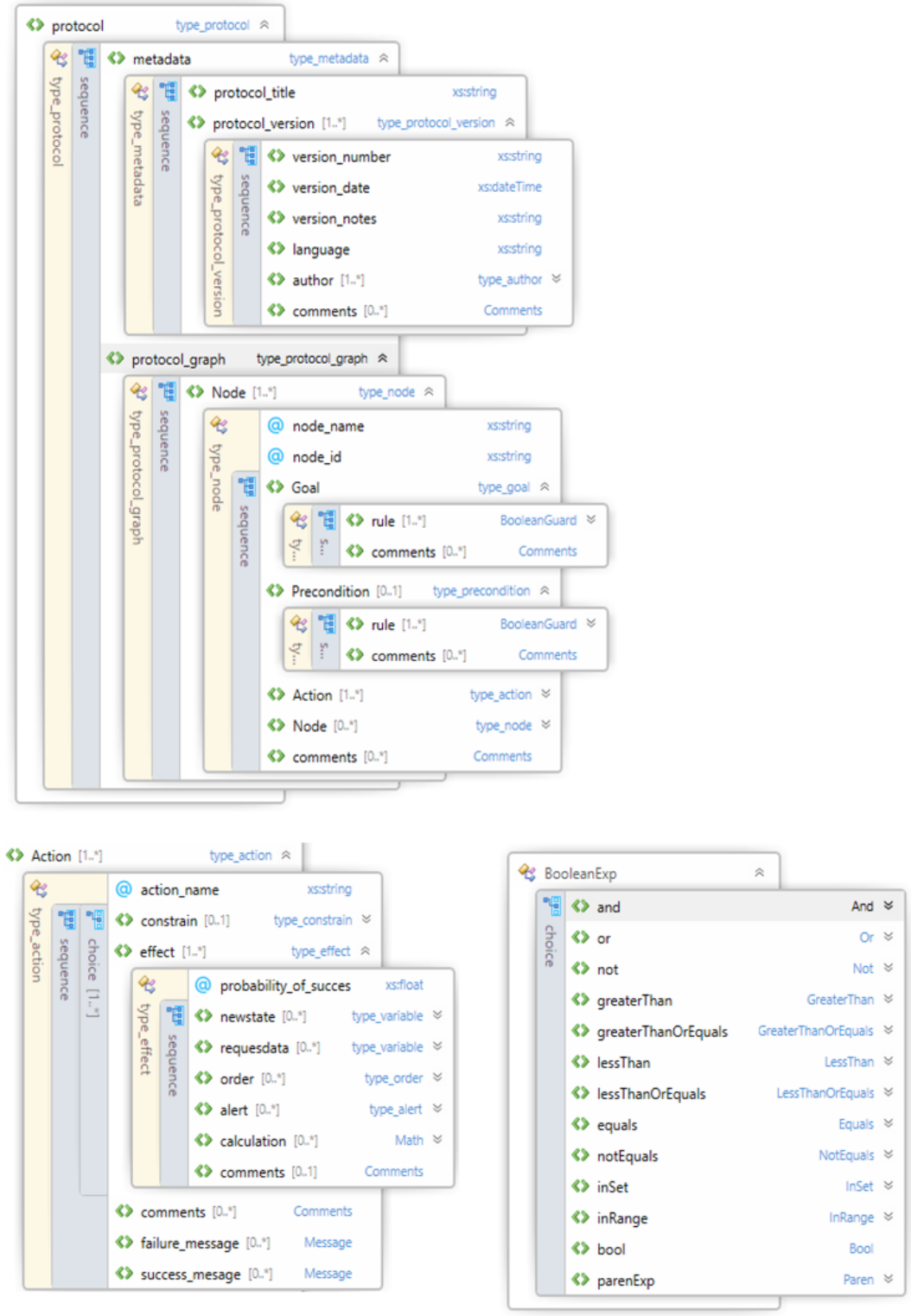
Fifth, the Paren type represents a parenthesized expression. It contains both the opening and closing parentheses and the expression within the parentheses.

To demonstrate the proposed knowledge representation, we will now show the code of an XML example for the medical situation of choosing a treatment when bacteria are present in the urine. In our example, knowledge is used to suggest a treatment when bacteria are present and their presence is captured in the record. The goal of our example is a patient free of bacteria in the urine. When this is conceptualized as a state machine system, the goal will be achieved by a state machine node representing the patient no longer having the bacteria present in the urine. Our preconditions will be satisfied and rules will be triggered by the colony-forming unit (CFU) value found in a previous urine culture test. There can be multiple kinds of potential actions. We want to address two common actions, a drug order for treatment according to the CFU count and an alert to the physician to determine if there remain bacteria in the urine.

If the patient's health improves but the lab result indicates there is still the presence of bacteria in the urine, but with a lower CFU, the system should evaluate the conditions and suggest a different treatment more matched to the patient conditions. The goal will still be the same, since the presence of the bacteria is still positive, but the preconditions (trigger) and the actions will be different.

For illustrative purposes, Multimedia Appendix 1 shows the contrast of an Arden medical logic module (MLM) for this task and our XML file representation, both with the same logic of treating a patient with bacterial infection. They share some similarities, as intended in the design structure. In the maintenance section, it is clear that the Arden syntax allows more details, being able to hold information for the guideline's author, version, institution, validation, and more. Both contain the same encoded statements (medical knowledge), in this case CFUs over 100,000, which trigger the generation of an antibiotic order. A big difference is the existence of the encoded goal in our knowledge representation proposal, where its existence with our state machine behavior makes actions no longer dependent only on the goal itself but also on the current state of the patient. This means that if the same goal is supplied for different patient states, it can lead to different actions, like different treatments or a different message sent to the physician.

Figure 3. XML Schema Definition file schema.



Discussion

Summary of Findings

In Table 1, we presented a comparison between some common CIGs used for CDSS. Each one has a unique representation language used to encode the medical knowledge. Our approach uses XML in order to take advantage of its flexible format to facilitate the electronic sharing of structured data. More importantly, we can take advantage of its flexibility to standardize CIG structure design in a low-specification language that is ubiquitous, universal, and human readable.

Our plans involve the computerization of antenatal health care guidelines with our knowledge representation for a web-based CDSS, and then a mobile app. The clinical decisions suggested will assist clinicians in three situations. First, it will evaluate obstetrical risk and inform clinicians and raise health alerts about patients' medical conditions, like pre-eclampsia, diabetes,

or premature birth. Second, it will suggest medical procedures and drugs according to the gestational age of the mother and her clinical condition. Third, it will make referral suggestions for transferring patients from primary to secondary care and vice versa.

As discussed in the Introduction, our knowledge representation was designed for an antenatal care CDSS project with requirements for both a web and mobile CDSS but also designed to be usable for many other purposes as part of future implementations of other clinical applications. Future applications will test the adequacy of the framework to support additional information and workflow needs of CPGs for other health services and allow us to expand the framework as needed.

For our particular antenatal domain problem scenarios, we found our current framework and implementation methods adequate to capture the workflow and decision logic of the existing guidelines (Table 2). The domain knowledge, task knowledge, and inferences required were easily extracted and separable into the architecture components and the underlying representation. We did not find it necessary to use a more complex representation, such as the alternatives discussed earlier in the paper (Table 1). Adequate domain object representations were accommodated in the framework as well as the representation of the task ontology we used. The inferences required fit into the simple rule syntax and the deterministic state change representations in the state machine conceptualization. The inferences could also easily be separated in the modular structure of the described state machine, both hierarchically and by arranging rule application precedence. For the domain ontology and task ontology in particular, we did not need a more complex hierarchical ontology, such as those in published biomedical ontologies. Semantic elements such as object properties and instances were captured in a straightforward manner in our XML framework.

It is possible to accommodate either controlled vocabulary terms from standardized sources in the XML representation or ad hoc terms. We anticipate that controlled vocabulary terms will be needed for more complex systems for interoperability with other care systems or to accommodate a greater variety of users than our current system allows. Standardized terms would also result in a more understandable human-computer interface. In more complex systems, there is often a need to include common terms for the purpose of human and machine reference and communication, but with minimal addition, the current approach appears to accommodate any Arden syntax medical logic module representation.

The task modeling ontology we chose was also adequate for our current system and easily represented in XML. The elements of the state machine process representation for the system's problem-solving steps that achieve a task goal and decompose tasks into subtasks were adequate for the antenatal system. A stochastic process representation addition is anticipated for other domains but can be accompanied within a stochastic state machine framework. The rule representation used also proved adequate for capturing the performance of Arden syntax MLMs.

An important step to consider in future projects will be parsing existing Arden syntax MLMs to our representations, since Arden syntax is an official and commonly used standard and there are plenty of MLMs with encoded medical knowledge using it. Certainly, at its current stage of development, our approach is not able to capture the more sophisticated logic required by the axioms using semantic web rules, such as Semantic Web Rule Language [34], which is commonly used in Web Ontology Language–based CDSSs.

We have sacrificed generality in the representation to achieve a minimal design footprint and execution efficiency across multiple computing environments while adequately capturing the guidelines required for our project. It is possible to think of the current representation as a compiled or low-level interpreted version of a more complex representation of guidelines that retains the computation resource efficiency and probability of the current approach for both web applications and mobile apps. In that case, an associated higher-level representation allowing for more generality and a friendlier user development environment based on the more complex representations would be possible.

In the future, having an interface design tool for authoring and editing our XML-represented protocols will be useful to ease the encoding of knowledge. XML is not hard to understand once people get used to it, but this does not mean that there cannot be another tool to encode medical knowledge more easily with XML. A graphic user interface that is object oriented can likely be designed to help the encoding process for health care professionals, who are more motivated than computer engineers to participate in encoding medical knowledge.

Conclusions

Usability of a CDSS depends heavily on the match of system flow to health care workflow. Allowing for consistent stepwise processing of health data over time can support adherence to best clinical practice. Consistent with best-practice CPGs, the intent of the CDSS is to reduce the caregiver's mental load and prevent possible errors in clinical tasks that involve the analysis of a patient's status and the use of this context for action decisions. Our knowledge representation framework incorporates fundamental elements of other CIGs used in CDSSs in medicine to encode a number of antenatal health care CPGs and associated clinical workflows. The framework appears general enough to be useful with other CPG-to-CIG projects in medicine.

XML proved to be a language expressive enough to describe the planning problems in a computable form and both restrictive and expressive enough to implement in a clinical system. It can be effective for mobile apps, where intermittent communication requires a small-footprint autonomous app. It can be used to incorporate overlapping capabilities of more specialized CIGs in medicine. These qualities of the XML language give it viability for use in CDSSs as a knowledge engine core that is based on a widely available and understood collection of technologies for web applications and mobile apps.

Acknowledgments

This work was supported by Departamento Nacional de Planeación through the Desarrollo de Soluciones en CTi para Telesalud en el Departamento de Antioquia project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example: comparison between Arden syntax and our XML file in starting antibiotic treatment.
[DOCX File , 18 KB - formative_v4i10e17512_app1.docx]

References

1. Grupo Desarrollador de la Guía - Universidad Nacional de Colombia. Guía de práctica clínica para la prevención, detección temprana y tratamiento de las complicaciones del embarazo, parto y puerperio: sección 3. Infecciones en el embarazo: ruptura prematura de membranas (RPM). Rev Colomb Obstet Ginecol 2015 Dec 21;66(4):263. [doi: 10.18597/rcog.293]
2. Berner E. Clinical decision support systems: state of the art. Agency for Healthcare Research and Quality. 2009. URL: https://digital.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf [accessed 2020-09-24]
3. Tenório JM, Hummel AD, Cohrs FM, Sdepanian VL, Pisa IT, de Fátima Marin H. Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease. Int J Med Inform 2011 Nov;80(11):793-802 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.08.001] [Medline: 21917512]
4. Peleg M. Computer-interpretable clinical guidelines: a methodological review. J Biomed Inform 2013 Aug;46(4):744-763 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.009] [Medline: 23806274]
5. Keiffer MR. Utilization of clinical practice guidelines: barriers and facilitators. Nurs Clin North Am 2015 Jun;50(2):327-345. [doi: 10.1016/j.cnur.2015.03.007] [Medline: 25999074]
6. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
7. Goud R, Hasman A, Peek N. Development of a guideline-based decision support system with explanation facilities for outpatient therapy. Comput Methods Programs Biomed 2008 Aug;91(2):145-153. [doi: 10.1016/j.cmpb.2008.03.006] [Medline: 18490078]
8. Amoakoh-Coleman M, Borgstein AB, Sondaal SF, Grobbee DE, Miltenburg AS, Verwijs M, et al. Effectiveness of mHealth Interventions Targeting Health Care Workers to Improve Pregnancy Outcomes in Low- and Middle-Income Countries: A Systematic Review. J Med Internet Res 2016 Aug 19;18(8):e226 [FREE Full text] [doi: 10.2196/jmir.5533] [Medline: 27543152]
9. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the Next-Generation Perinatal Care: An Overview of the Literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
10. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and Appropriateness of mHealth Interventions for Maternal and Child Health: Systematic Review. JMIR Mhealth Uhealth 2018 Jan 09;6(1):e7 [FREE Full text] [doi: 10.2196/mhealth.8998] [Medline: 29317380]
11. Bray T, Paoli J, Sperberg-McQueen C, Maler E, Yergeau F. Extensible Markup Language (XML) 1.0 (Fifth Edition). World Wide Web Consortium (W3C). 2008 Nov 26. URL: https://www.w3.org/TR/REC-xml/ [accessed 2020-09-29]
12. Zunke S, D'Souza V. JSON vs XML: A Comparative Performance Analysis of Data Exchange Formats. Int J Comput Sci Netw 2014;3(4):257-261 [FREE Full text]
13. Zolhavarieh S, Parry D, Bai Q. Issues Associated With the Use of Semantic Web Technology in Knowledge Acquisition for Clinical Decision Support Systems: Systematic Review of the Literature. JMIR Med Inform 2017 Jul 05;5(3):e18 [FREE Full text] [doi: 10.2196/medinform.6169] [Medline: 28679487]
14. Jing X, Hardiker NR, Kay S, Gao Y. Identifying Principles for the Construction of an Ontology-Based Knowledge Base: A Case Study Approach. JMIR Med Inform 2018 Dec 21;6(4):e52 [FREE Full text] [doi: 10.2196/medinform.9979] [Medline: 30578220]
15. Teodoro D, Pasche E, Gobeill J, Emonet S, Ruch P, Lovis C. Building a transnational biosurveillance network using semantic web technologies: requirements, design, and preliminary evaluation. J Med Internet Res 2012;14(3):e73 [FREE Full text] [doi: 10.2196/jmir.2043] [Medline: 22642960]
16. Jeannot J, Scherer F, Pittet V, Burnand B, Vader J. Use of the World Wide Web to implement clinical practice guidelines: a feasibility study. J Med Internet Res 2003;5(2):e12 [FREE Full text] [doi: 10.2196/jmir.5.2.e12] [Medline: 12857668]

17. Bachmann KF, Vetter C, Wenzel L, Konrad C, Vogt AP. Implementation and Evaluation of a Web-Based Distribution System For Anesthesia Department Guidelines and Standard Operating Procedures: Qualitative Study and Content Analysis. J Med Internet Res 2019 Aug 15;21(8):e14482 [FREE Full text] [doi: 10.2196/14482] [Medline: 31418427]
18. Falkman G, Gustafsson M, Jontell M, Torgersson O. SOMWeb: a semantic web-based system for supporting collaboration of distributed medical communities of practice. J Med Internet Res 2008 Aug 26;10(3):e25 [FREE Full text] [doi: 10.2196/jmir.1059] [Medline: 18725355]
19. Microsoft. ASP.NET MVC Overview. Microsoft Docs. 2009 Jan 27. URL: https://docs.microsoft.com/en-us/aspnet/mvc/overview/older-versions-1/overview/asp-net-mvc-overview [accessed 2020-09-29]
20. OpenClinical workshop: Methods for representing computer-interpretable clinical guidelines. Open Clinical. URL: http://www.openclinical.org/gmmworkshop2001.html [accessed 2017-08-22]
21. De Clercq P, Kaiser K, Hasman A. Computer-Interpretable Guideline formalisms. Stud Health Technol Inform 2008;139:22-43 [FREE Full text] [Medline: 18806319]
22. de Clercq PA, Blom JA, Korsten HHM, Hasman A. Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif Intell Med 2004 May;31(1):1-27. [doi: 10.1016/j.artmed.2004.02.003] [Medline: 15182844]
23. Mulyar N, van der Aalst WMP, Peleg M. A pattern-based analysis of clinical computer-interpretable guideline modeling languages. J Am Med Inform Assoc 2007;14(6):781-787 [FREE Full text] [doi: 10.1197/jamia.M2389] [Medline: 17712087]
24. Wang D, Peleg M, Tu SW, Boxwala AA, Ogunyemi O, Zeng Q, et al. Design and implementation of the GLIF3 guideline execution engine. J Biomed Inform 2004 Oct;37(5):305-318 [FREE Full text] [doi: 10.1016/j.jbi.2004.06.002] [Medline: 15488745]
25. Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, et al. Comparing computer-interpretable guideline models: a case-study approach. J Am Med Inform Assoc 2003;10(1):52-68 [FREE Full text] [doi: 10.1197/jamia.m1135] [Medline: 12509357]
26. Vargas-Vera M, Motta E, Domingue J, Buckingham Shum S, Lanzoni M. Knowledge extraction by using an ontology-based annotation tool. 2001 Presented at: Proceedings of the First International Conference on Knowledge Capture: K-CAP01; Oct 21-23, 2001; Victoria, BC, Canada.
27. Yu C, Shen B. XML, Ontologies, and Their Clinical Applications. Adv Exp Med Biol 2016;939:259-287. [doi: 10.1007/978-981-10-1503-8_11] [Medline: 27807751]
28. Pronovost PJ. Enhancing physicians' use of clinical guidelines. JAMA 2013 Dec 18;310(23):2501-2502. [doi: 10.1001/jama.2013.281334] [Medline: 24310916]
29. Ministerio de Salud y Protección Social. Guía de práctica clínica para el diagnóstico, tratamiento y seguimiento de la diabetes Gestacional. Bogotá, Colombia: Ministerio de Salud y Protección Social; Nov 2015.
30. Secretaría Distrital de Salud de Bogotá, Asociación Bogotana de Obstetricia y Ginecología (Asbog). Guía de control prenatal y factores de riesgo. 2013. URL: http://www.saludcapital.gov.co/DDS/Publicaciones/GUIA%201.%20%20CONTROL%20PRENATAL%20Y%20FACTORES%20DE%20RIESGO.pdf [accessed 2020-09-24]
31. Grupo de trabajo de la Guía de práctica clínica de atención en el embarazo y puerperio, Ministerio de Sanidad, Servicios Sociales e Igualdad. Guía de práctica clínica de atención en el embarazo y puerperio. 2010. URL: https://portal.guiasalud.es/wp-content/uploads/2018/12/GPC_533_Embarazo_AETSA_compl.pdf [accessed 2020-09-29]
32. Ministerio de Salud y la Protección Social, Fondo de Poblaciones de las Naciones Unidas. Guía de práctica clínica (GPC) basada en la evidencia para la atención integral de la sífilis gestacional y congénita. 2014. URL: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/IETS/gpc%20%E2%80%93guia-completa-sifilis.pdf [accessed 2020-09-24]
33. Boxwala AA, Rocha BH, Maviglia S, Kashyap V, Meltzer S, Kim J, et al. A multi-layered framework for disseminating knowledge for computer-based decision support. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i132-i139 [FREE Full text] [doi: 10.1136/amiajnl-2011-000334] [Medline: 22052898]
34. O'Connor M, Shankar R, Nyulas C, Tu S, Das A. Developing a Web-Based Application using OWL and SWRL. 2008 Presented at: AAAI 2008 Spring Symposium: AI Meets Business Rules and Process Management; March 26-28, 2008; Stanford, CA.

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
CFU: colony-forming unit
CIG: computer-interpretable guidelines
CPG: clinical practice guideline
FCM: fuzzy cognitive map
MLM: medical logic module
MVC: model-view-controller
XSD: XML Schema Definition

Edited by C Lovis; submitted 17.12.19; this is a non–peer-reviewed article;accepted 16.08.20; published 16.10.20.

Please cite as:
Torres Silva EA, Uribe S, Smith J, Luna Gomez IF, Florez-Arango JF
XML Data and Knowledge-Encoding Structure for a Web-Based and Mobile Antenatal Clinical Decision Support System: Development Study
JMIR Form Res 2020;4(10):e17512
URL: http://formative.jmir.org/2020/10/e17512/
doi:10.2196/17512
PMID:33064087

Original Paper

Coding Systems for Clinical Decision Support: Theoretical and Real-World Comparative Analysis

Nicolas Delvaux[1], MD, PhD; Bert Vaes[1], MD, PhD; Bert Aertgeerts[1], MD, PhD; Stijn Van de Velde[1,2], PhD; Robert Vander Stichele[3], MD, PhD; Peter Nyberg[4], MD, PhD; Mieke Vermandere[1], PhD, MD

[1]Department of Public Health and Primary Care, Katholieke Universiteit Leuven, Leuven, Belgium

[2]Division for Health Services, Norwegian Institute of Public Health, Oslo, Norway

[3]Department of Medical Informatics, Ghent University, Ghent, Belgium

[4]Duodecim Publishing Company Ltd, Helsinki, Finland

Corresponding Author:
Nicolas Delvaux, MD, PhD
Department of Public Health and Primary Care
Katholieke Universiteit Leuven
Kapucijnenvoer 33 Blok J - box 7001
Leuven, B-3000
Belgium
Phone: 32 16379885
Email: nicolas.delvaux@kuleuven.be

Abstract

Background: Effective clinical decision support systems require accurate translation of practice recommendations into machine-readable artifacts; developing code sets that represent clinical concepts are an important step in this process. Many clinical coding systems are currently used in electronic health records, and it is unclear whether all of these systems are capable of efficiently representing the clinical concepts required in executing clinical decision support systems.

Objective: The aim of this study was to evaluate which clinical coding systems are capable of efficiently representing clinical concepts that are necessary for translating artifacts into executable code for clinical decision support systems.

Methods: Two methods were used to evaluate a set of clinical coding systems. In a theoretical approach, we extracted all the clinical concepts from 3 preventive care recommendations and constructed a series of code sets containing codes from a single clinical coding system. In a practical approach using data from a real-world setting, we studied the content of 1890 code sets used in an internationally available clinical decision support system and compared the usage of various clinical coding systems.

Results: SNOMED CT and ICD-10 (International Classification of Diseases, Tenth Revision) proved to be the most accurate clinical coding systems for most concepts in our theoretical evaluation. In our practical evaluation, we found that International Classification of Diseases (Tenth Revision) was most often used to construct code sets. Some coding systems were very accurate in representing specific types of clinical concepts, for example, LOINC (Logical Observation Identifiers Names and Codes) for investigation results and ATC (Anatomical Therapeutic Chemical Classification) for drugs.

Conclusions: No single coding system seems to fulfill all the needs for representing clinical concepts for clinical decision support systems. Comprehensiveness of the coding systems seems to be offset by complexity and forms a barrier to usability for code set construction. Clinical vocabularies mapped to multiple clinical coding systems could facilitate clinical code set construction.

(JMIR Form Res 2020;4(10):e16094) doi:10.2196/16094

KEYWORDS

clinical decision support systems; clinical coding; medical informatics; electronic health records

Introduction

Clinical decision support systems are considered to be an important vehicle for implementing new evidence and knowledge into daily practice [1,2]. Effective health care implies well-informed choices and decisions based on reliable evidence, with attention to individual needs and drawn from clinical experience [3]. Despite demonstrating rather small effects on

adherence in clinical trials [4], clinical decision support systems are widely accepted as an important strategy for knowledge translation [5]. Clinical decision support systems generate patient-specific recommendations by matching individual patient characteristics to a knowledge base [6]; they are available in various formats and presentations, but an important variant of clinical decision support systems is guideline-driven, generating reminders based on formal rules and algorithms. The 3 essential components of a clinical decision support systems are (1) a knowledge base, (2) an inference or reasoning engine, and (3) an interface that can communicate with the user [7]. The knowledge base of a clinical decision support system consists of clinical practice recommendations that have been translated into machine-readable algorithms or artifacts. Artifacts are formal expressions of the recommendations in clinical guidelines. They include concepts from many different aspects from clinical practice, such as diagnoses, procedures, observations, or drugs. For the inference engine to be able to query the database of an electronic health record, each concept needs to be translated into a set of clinical codes also known as a clinical code set [8]. Collections of clinical codes, or clinical coding systems, are currently in use by electronic health records to represent clinical concepts, all with different finalities and purposes. Table 1 illustrates several of the current clinical coding systems used in electronic health records. Classifications include a form of taxonomy or structure of the included codes [8]. In some cases, this taxonomy is basic, such as those in the World Health Organization (WHO) family of classifications. For instance, in the International Classification of Diseases Tenth Revision (ICD-10), the structure is reflected by hierarchical alphanumeric codes. For instance, in ICD-10, the code for calculus of the kidney (N20.0) is a child of calculus of the kidney or ureter (N20) which is in turn a child of diseases of the genitourinary system (N). These types of taxonomies prevent clinical concepts from existing more than once in the classification but sometimes simplify more complicated concepts, such as pulmonary infections that could be classified as a pulmonary disease but also as an infectious disease. More complicated relationships are possible in ontologies, such as that in SNOMED CT, which includes not only “is a” hierarchy but also “has finding site” or “has causative agent [9].”

Table 1. Overview of some clinical terminologies used in electronic health records including their domain coverage and purpose. This list is not exhaustive.

Clinical coding system	Domain coverage	Type, purpose
SNOMED CT	Multiple areas (diagnoses, allergies, symptoms, etc)	Terminology, clinical documentation
International Classification of Diseases (ICD)	Diagnoses, some procedures	Classification, reporting
Current Procedural Terminology (CPT)	Procedures	Terminology, clinical documentation
Logical Observation Identifiers Names and Codes (LOINC)	Laboratory tests	Terminology, clinical documentation
Anatomical Therapeutic Chemical (ATC)	Drugs	Classification, reporting
International Classification of Primary Care (ICPC)	Diagnoses, reasons for encounter, some procedures	Classification, reporting

Most attention, when evaluating clinical decision support systems, is directed at ensuring technical interoperability and digitally structured data within the electronic health record as these determine the appropriateness of clinical decision support system alerts. Electronic health records currently use a variety of clinical coding systems to structure and represent clinical data, often with different purposes. Despite reports [8] on methods to translate clinical practice recommendations into interoperable artifacts and their code sets, it remains unclear whether currently used clinical coding systems are capable of representing the concepts needed for clinical decision support systems. Designers of terminology for clinical decision support systems have suggested that no one clinical coding system is capable of describing all necessary clinical concepts and that concurrent use of multiple terminologies is required [10]. The aim of this study was to evaluate whether currently used clinical coding systems are capable of efficiently representing the clinical concepts that are required for translating artifacts into executable code for clinical decision support systems.

Methods

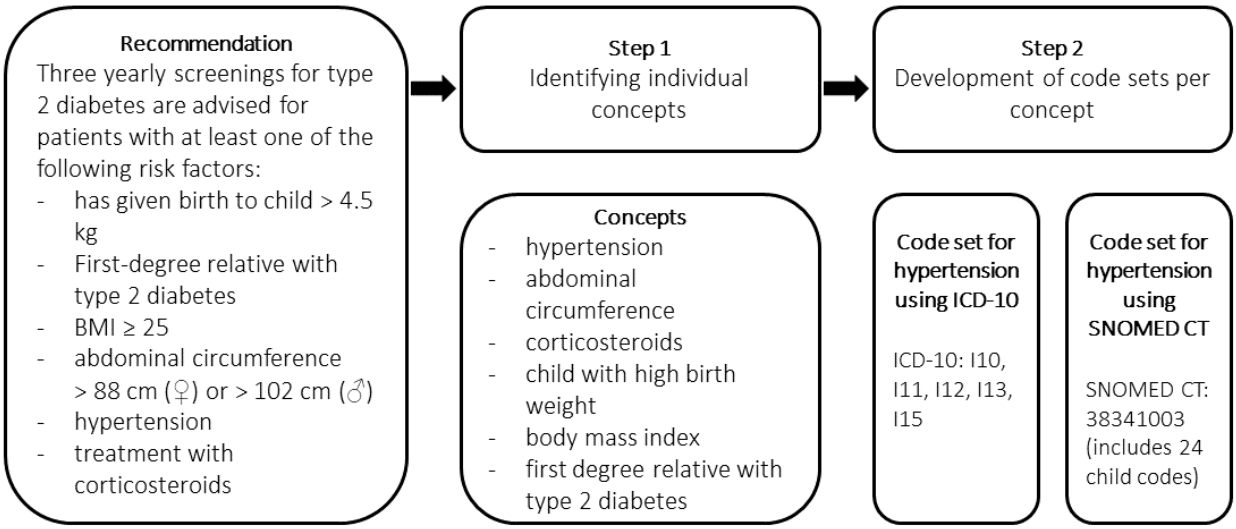
We used 2 separate methods—(1) theoretical and (2) practical evaluation using data from a real-world setting.

Theoretical Evaluation

We aimed to evaluate whether a selection of clinical coding systems was capable of representing the clinical concepts in a small set of recommendations and how many codes were required for this. First, we designed clinical decision support system artifacts based on 3 recommendations for preventive care which included concepts relevant to primary care [11]. We chose these recommendations because they were evidence-based and locally applicable, in addition, adherence to these recommendations was suboptimal. The recommendations used for this evaluation are described in Multimedia Appendix 1. We identified the clinical information described in the recommendations and isolated all individual clinical concepts to be used in the artifact. For each clinical concept, a clinical code set was constructed containing codes from a single system. We repeated this task for a selection of classifications, terminologies, and coding systems, including International Classification of Primary Care (ICPC)–2, International

Classification of Diseases Tenth Revision (ICD-10) –Clinical Modification (-CM), SNOMED CT, Anatomical Therapeutic Chemical (ATC) Classification, and Logical Observation Identifiers Names and Codes (LOINC). An example of this process is illustrated in Figure 1.

Figure 1. Process for the development of code sets for a guideline recommendation. BMI: body mass index; ICD-10: International Classification of Diseases, Tenth Revision.



However, not all coding systems were capable of representing the clinical concepts in an artifact. For example, for *acute exacerbation of chronic obstructive pulmonary disease*, there is no ICPC-2 code. The ICPC classification only includes a code for *chronic obstructive pulmonary disease* (R95) but does not allow specification of an acute exacerbation. The ICD-10 classification contains 2 codes for an acute exacerbation of chronic obstructive pulmonary disease: *chronic obstructive pulmonary disease with acute lower respiratory infection* (J44.0) and *chronic obstructive pulmonary disease with acute exacerbation, unspecified* (J44.1). Together, the codes J44.0 and J44.1 constitute the ICD-10 code set for *acute exacerbation of chronic obstructive pulmonary disease*. If the constructed code set was incapable of fully representing the clinical concept due insufficient granularity or overlap with other concepts, the code set was excluded from further evaluation. For SNOMED CT, we included all the codes, including child codes, that were required to fully represent the clinical concept.

Practical Evaluation

Rather than constructing new code sets from recommendations, we studied the content of a large database of existing code sets used in an internationally available clinical decision support system, the Evidence-Based Medicine electronic Decision Support (EBMeDS, Duodecim Medical Publications Ltd). At the time of this study, EBMeDS contained 1890 concepts, and for each concept, a code set had been constructed using a large number of clinical coding systems currently in use. In addition to international coding systems, code sets also included local or national clinical coding systems and, in some cases, even electronic health record–specific proprietary coding systems. For each code set, we reviewed all included clinical coding systems and compared their usage. Clinical coding systems that were used in less than 3% of the code sets were not reported as these were always electronic health record proprietary coding systems.

Results

Theoretical Evaluation

For the 3 recommendations, we identified 21 different clinical concepts which we defined using 5 different clinical coding systems (see Multimedia Appendix 2 for the full code sets). Of these concepts, 10 were diagnoses, 3 were procedures, 2 were risk factors, 2 were laboratory tests, 2 were drugs, and 2 were vaccines. Table 2 show the distribution of the clinical coding systems for which we were able to create code sets describing the clinical concepts. In the case of the diagnosis concepts, for 9 out of 10 concepts, we were able to create a code set of SNOMED CT codes, for 8 out of 10 concepts we were able to create a code set of ICD-10 codes, and for 2 out of 10 concepts, we were able to create a code set of ICPC-2 codes. We did not find SNOMED CT codes for the concept *having given birth to a child over 4.5 kg*; we did not find ICD-10 codes for *presence of a cochlear implant* and *stress hyperglycemia*; and we did not find ICPC-2 codes for *asplenia*, *sickle cell disease*, *hemoglobinopathy*, *cerebrospinal fluid leak*, *presence of a cochlear implant*, *weak immunity*, *stress hyperglycemia*, and *having given birth to a child over 4.5 kg*. The SNOMED CT code sets included a median of 11 (range 1-219) codes, whereas the ICD-10 code sets included a median of 1 (range 1-3) codes.

Table 2. Code sets identified from 3 preventive care recommendations.

Concept type and coding system	Code sets per concept type[a], n (%)	Codes per set, median (range[b])
Diagnosis (n=10)		
SNOMED CT[c]	9 (90)	11 (1-219)
ICD-10(-CM)[d]	8 (80)	1 (1-3)
ICPC-2[e]	2 2(0)	1.5 (1-2)
Drugs (n=2)		
ATC[f]	1 (50)	1 (—[g])
SNOMED CT	2 (100)	108 (53-163)
Vaccines (n=2)		
ATC	2 (100)	1 (—)
SNOMED CT	2 (100)	3 (1-5)
Risk factor (n=2)		
ICPC-2	2 (100)	1 (—)
ICD-10(-CM)	2 (100)	1 (—)
SNOMED CT	2 (100)	9 (9-9)
Procedure (n=3)		
ICD-10(-CM)	2 (67)	1.5 (1-2)
SNOMED CT	3 (100)	1 (1-37)
LOINC[h]	3 (100)	5 (2-9)
Investigation results (n=2)		
SNOMED CT	1 (50)	2 (—)
LOINC	2 (100)	8 (7-9)

[a]The proportion of code sets per total number of concepts represents the proportion of clinical concepts for which a set of codes was found that matched the clinical concept.

[b]Minimum to maximum.

[c]SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms.

[d]ICD-10(-CM): International Classification of Diseases, Tenth Revision (–Clinical Modification).

[e]ICPC: International Classification of Primary Care.

[f]ATC: Anatomical Therapeutic Chemical Classification.

[g]Indicates that the range is defined by a single value.

[h]LOINC: Logical Observation Identifiers Names and Codes.

Practical Evaluation

The majority of the predefined EBMeDS concepts were diagnosis or drug concepts, with very few risk factor or vaccine concepts. Table 3 shows the distribution of clinical coding systems used in the code sets for the clinical concepts including the number of individual codes.

Table 3. Code sets defined in the EBMeDS service.

Concept type and coding system	Code sets per concept type[a], n (%)	Codes per set, median (range[b])
Diagnosis (n=790)		
International		
ICD-10[c]	739 (93.5)	1 (1-53)
ICPC-2[d]	223 (28.2)	1 (1-26)
ICD-9-CM[e]	93 (11.8)	1 (1-550)
SNOMED CT[f]	89 (11.3)	2 (1-36)
Drugs (n=556)		
International		
ATC[g]	481 (86.5)	1 (1-245)
SNOMED CT	25 (4.5)	1 (1-2)
National		
Read codes (United Kingdom)	17 (3.1)	1 (1-5)
Investigation results (n=317)		
International		
LOINC[h]	116 (36.6)	1 (1-8)
Nomenclature for Properties and Units	76 (24.0)	1 (1-9)
National		
KL Finnish classification for laboratory investigations	271 (85.5)	2 (1-60)
Read codes (United Kingdom)	25 (7.9)	1 (1-4)
Proprietary EHR[i]		
Meldola Hospital measurement classification (Italy)	54 (17.0)	1 (1-2)
SoSoeMe measurement classification (Belgium)	67 (21.1)	1 (1-3)
Health One measurement classification (Belgium)	63 (19.9)	1 (1-3)
Procedures (n=214)		
International		
SNOMED CT	8 (3.7)	1 (1-43)
ICD-9-CM	27 (12.6)	1 (1-16)
National		
Current Procedural Terminology	17 (7.9)	7 (1-42)
Nordic procedure codes	179 (83.6)	1 (1-468)
Proprietary EHR		
Quantros Organization (United States)	7 (3.3)	1 (—[j])
Risk factors (n=2)		
International		
ICD-10	1 (50.0)	1 (—)
Proprietary EHR		
Health One measurement classification (Belgium)	1 (50.0)	1 (—)
Vaccines (n=11)		
International		
ATC	11 (100.0)	1 (1-12)

Concept type and coding system	Code sets per concept type[a], n (%)	Codes per set, median (range[b])
National		
ROKVALM Finnish vaccination codes	9 (90.9)	7.5 (1-19)
ROK Finnish vaccination codes	8 (81.8)	1 (1-7)

[a]The proportion of code sets per total number of concepts represents the proportion of clinical concepts for which a set of codes was found that matched the clinical concept.

[b]Minimum to maximum.

[c]ICD-10: International Classification of Diseases, Tenth Revision.

[d]ICPC: International Classification of Primary Care.

[e]ICD-9-CM: International Classification of Diseases, Ninth Revision–Clinical Modification.

[f]SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms.

[g]ATC: Anatomical Therapeutic Chemical Classification.

[h]LOINC: Logical Observation Identifiers Names and Codes.

[i]EHR: electronic health record.

[j]Indicates that the range is defined by a single value.

Discussion

Principal Findings

In the theoretical evaluation of clinical coding systems, we found that SNOMED CT and ICD-10 were capable of describing the majority of diagnosis concepts; however, for some clinical concepts, a very large number of codes was required. In our theoretical analysis, SNOMED CT was superior to ATC for drug concepts, which in our small set was entirely due to the fact that ATC does not always define a route of administration. The small sample of code sets for drug concepts probably exaggerated the superiority of SNOMED CT in comparison to ATC. The shortcomings of ATC—sometimes lacking a route of administration and not always including all substances of compound drugs—are probably less influential than our evaluation suggests. The procedure concepts in this study could all be mapped to SNOMED CT and LOINC. The only instances where SNOMED CT did not fully represent clinical concepts were investigation results. Code sets with SNOMED CT codes often included a lot more codes than those included for other coding systems, with a median of 108 codes for drug concepts. Constructing code sets with this number of codes may challenge the feasibility of this task. When compared to the results of the practical evaluation of current code sets in the EBMeDS database of clinical concepts, SNOMED CT was noticeably less present. Only 11.3% of diagnosis concepts (89/790) were mapped to SNOMED CT codes and even less for drug (25/556, 4.5%) and procedure concepts (8/214, 3.7%). Most used in EBMeDS were the ICD-10 and ICD-9 families, ICPC-2, ATC, and LOINC. For procedure, investigation result, and vaccine concepts, there appeared to be a lack of internationally accepted coding systems, since a lot of code sets included national coding systems or even electronic health record–specific proprietary coding systems.

There are several reasons for the popularity of the World Health Organization family of classifications. The comprehensiveness and widespread use of these classifications make them popular for clinical coding. The clinical codes in ICD are alphanumeric and arranged hierarchically. This allows for truncation of the codes in order to include a large number of child concepts with codes starting with the same sequence. For instance, by truncating the code K29*, it is possible to include the 10 different subcategories of gastritis and duodenitis without having to include each of these 10 codes. SNOMED CT, the other clinical coding system capable of representing a large majority of the clinical concepts, does not contain this feature because the unique identifiers of each code do not mirror the relationship to one another. This does not allow for truncation of the codes and explains why such a large number of codes are required to define each concept. As described earlier, ICD-10(-CM) only contains single parent-child relationships, but SNOMED CT includes multiple relationships. In investigating further, we found that if a clinical decision support system could recognize all possible SNOMED CT hierarchical relationships through a programmed expression, then the number of codes required to identify a concept was similar for SNOMED CT and ICD-10(-CM) (Multimedia Appendix 2). If, however, the clinical decision support system was only capable of recognizing the concept's unique SNOMED CT identifier without its relationships, then a much larger number of codes was necessary. To date, very few SNOMED CT codes are included in EBMeDS mappings, and most mappings are for demonstration purposes only because the license to use SNOMED CT in EBMeDS has only recently been obtained. Therefore, the limited use of SNOMED CT may well be a consequence of fragmented uptake of this terminology in electronic health record systems that have integrated EBMeDS. ICPC-2, often used for documenting diseases and reasons for encounter in primary care, is of limited use in defining concepts required in clinical decision support systems; often the concepts defined in ICPC-2 are too broad and insufficiently detailed.

SNOMED CT has high sensitivity and specificity in representing clinical concepts [12,13]. However, creating code sets using SNOMED CT poses some important challenges. Through its poly-hierarchical structure, SNOMED CT creates an intricate web of clinical terms with multiple types of relationships defined through attributes. As opposed to a mono-hierarchical classification which has a branched structure, SNOMED CT has a profoundly complex web-like structure. This complexity

may be a barrier to implementing this clinical coding system. In addition, SNOMED CT contains a very large number of terms, which makes it very difficult to create clinical code sets [8].

LOINC is very good at defining laboratory and physiological tests, but similar to that of SNOMED CT, the complex and granular classification structure is problematic. Despite the capacity of SNOMED CT to describe procedure codes, many countries use their own proprietary clinical coding systems. This may be as a result of locally used procedure lists generated for billing purposes which may not be internationally applicable. National or electronic health record system proprietary clinical coding systems may be very useful for some aggregate use of the electronic health record, but a myriad of coding systems for small-scale use requires multiple mappings and increases the odds of inappropriate mapping.

Limitations

For this study, we chose to limit the number of recommendations that we analyzed for each of the coding systems or terminologies. Manual searching of codes and terms that applied to individual clinical concepts proved time consuming and is, therefore, not feasible for a larger set of recommendations. We also limited the clinical coding systems to those systems currently in use in Europe. Therefore, several clinical coding systems that were included in the practical evaluation were not included in the theoretical evaluation, such as Current Procedural Terminology. The code sets were constructed by one person (ND) and were not validated by a second reviewer. The small scale of this first assessment and the lack of external validation of the code sets demands caution when drawing conclusions, but some trends were clear. We correlated these trends with a large EBMeDS database of existing mappings used in an internationally available clinical decision support system. The use of clinical coding systems in EBMeDS may not necessarily imply that they are well suited for defining concepts but may merely mirror the de facto use of these coding systems in the electronic health records where EBMeDS is integrated.

Our study did not assess all possible domains that may need translation into clinical coding. We did not study any terminologies or classifications that attempt to structure concepts such as pain, distress, anxiety, or other more complicated concepts. These types of clinical information are currently often lacking in clinical decision support systems and remain underexposed in studies.

In addition, the findings from this study are limited to one particular aspect of clinical decision support systems, namely the efficiency of particular clinical coding systems in correctly defining clinical concepts required to translate recommendations in decision support rules. Our study does not evaluate the efficiency of these systems in assisting clinicians in high-quality documentation at the point of care. More important than the capacity of a clinical coding system to correctly define a clinical concept may be the capacity of clinicians to correctly use these systems to document clinical data into the electronic health record. In a study [14] on clinical decision support systems using gastrointestinal risk scores, when confronted with identical patients consulting for identical problems, differences in how clinical information was recorded led to almost 80% of inaccurate recommendations by the clinical decision support systems. Similarly, a recent study [15] on clinical decision support system alerts on potential adverse drug events showed that almost 9 out of 10 alerts were overridden and that more than 8 out of 10 of these overrides were appropriate [15]. Quite often, alerts were triggered on drugs that had been stopped but were inadequately documented in the electronic health record. These findings suggest that the true bottleneck in data quality is probably not due to limitations in data coding or terminologies but to the quality of the documentation by clinicians or other sources of bias in electronic health records [16].

Implications

One would expect that the more detailed a coding system becomes, the more suited it becomes for defining concepts necessary for clinical decision support systems. However, it is unclear whether SNOMED CT, currently the most comprehensive clinical coding system available, is also the best choice for developing clinical decision support system artifacts. Through its rich poly-hierarchical relationships, SNOMED CT is growing into a true ontology potentially allowing for consistent documentation of practically all aspects of health care [17]. A European comprehensive evaluation of SNOMED CT implementations recognized the pivotal role of SNOMED CT as a core reference terminology but placed it as a part of a greater ecosystem of terminologies [18]. Important advantages of SNOMED CT are its single ownership, unique source, and clear ontology-based architecture, including the capacity to postcoordinate (combine concepts to create new, more detailed concepts). This potential is offset by important disadvantages such as its complexity and granularity, which require a comprehensive understanding of its structure before it can be used for knowledge representation in a clinical decision support system. Moreover, SNOMED CT still needs to prove its usability and user-friendliness as a clinical coding dictionary at the point of care, since its comprehensiveness may very well be a burden rather than an advantage as illustrated in the satirical paper by Richard Williams [19]. Hence, clinical coding systems suited for clinical documentation may not necessarily be the most adequate for information retrieval or other secondary use such as clinical decision support systems. A possible solution to this problem could be the development of local vocabularies that contain clinical terms which are mapped to multiple clinical coding systems, including reference terminologies such as SNOMED CT [10,20]. This would allow clinicians, and other potential users, to code clinical information using routinely used terms, simultaneously documenting the data in multiple structures. Depending on the type of aggregate use, clinical decision support system, quality-of-care indicator measurement, pay-for-performance schemes, or health policies, different clinical codes can be queried in the electronic health record.

Conclusions

Translating recommendations from clinical guidelines into artifacts for clinical decision support systems is an important step in implementing evidence-based health care. Not all clinical coding systems used in electronic health records for routine collection of clinical data are equally efficient in defining the

concepts in clinical decision support system artifacts. Research is needed to study whether the use of more comprehensive clinical coding systems such as SNOMED CT influences the appropriateness of clinical decision support system alerts.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preventive care recommendations.
[PDF File (Adobe PDF File), 109 KB - formative_v4i10e16094_app1.pdf]

Multimedia Appendix 2
Code sets.
[PDF File (Adobe PDF File), 874 KB - formative_v4i10e16094_app2.pdf]

References

1. The learning healthcare system: workshop summary. Institute of Medicine Roundtable on Evidence-Based Medicine. Washington: National Academies Press; 2007. URL: https://www.ncbi.nlm.nih.gov/books/n/nap11903/pdf/ [accessed 2020-10-01]
2. Shojania K, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J. The effects of on-screen, point of care computer reminders on processes and outcomes of care. Cochrane Database Syst Rev 2009 Jul 08(3):CD001096 [FREE Full text] [doi: 10.1002/14651858.CD001096.pub2] [Medline: 19588323]
3. Straus S, Glasziou P, Richardson W, Haynes R. Evidence-Based Medicine: How To Practice And Teach It 4th Edition. Edinburgh: Churchill Livingstone; 2011.
4. Kwan J, Lo L, Ferguson J, Goldberg H, Diaz-Martinez J, Tomlinson G, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. BMJ 2020 Sep 17;370:m3216. [doi: 10.1136/bmj.m3216] [Medline: 32943437]
5. Sarkar U, Samal L. How effective are clinical decision support systems? BMJ 2020 Sep 17;370:m3499. [doi: 10.1136/bmj.m3499] [Medline: 32943393]
6. Haynes RB, Wilczynski NL, Computerized Clinical Decision Support System (CCDSS) Systematic Review Team. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: methods of a decision-maker-researcher partnership systematic review. Implement Sci 2010 Feb 05;5:12 [FREE Full text] [doi: 10.1186/1748-5908-5-12] [Medline: 20181104]
7. Tan J, Sheps S. Health Decision Support Systems. Maryland: Aspen Publishers Inc; 1998.
8. Williams R, Kontopantelis E, Buchan I, Peek N. Clinical code set engineering for reusing EHR data for research: a review. J Biomed Inform 2017 Jun;70:1-13 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.010] [Medline: 28442434]
9. SNOMED CT Starter Guide Internet. International Health Terminology Standards Development Organisation (IHTSDO). 2017 Jul. URL: https://confluence.ihtsdotools.org/display/DOCSTART?preview=/28742871/47677485/doc_StarterGuide_Current-en-US_INT_20170728.pdf [accessed 2020-07-01]
10. Lin Y, Staes CJ, Shields DE, Kandula V, Welch BM, Kawamoto K. Design, Development, and Initial Evaluation of a Terminology for Clinical Decision Support and Electronic Clinical Quality Measurement. AMIA Annu Symp Proc 2015;2015:843-851 [FREE Full text] [Medline: 26958220]
11. Aerts L, Baeten R, Govaerts F, Barthelomeeusen E, Boonen S, De BK, et al. Gezondheidsgids 2019: Handleiding voor preventie in de huisartsenpraktijk. Antwerpen: Domus Medica vzw; 2019. URL: http://www.ggids.be [accessed 2020-10-01]
12. Al-Hablani B. The use of automated SNOMED CT clinical coding in clinical decision support systems for preventive care. Perspect Health Inf Manag 2017;14(Winter):1f [FREE Full text] [Medline: 28566995]
13. Odigie E, Lacson R, Raja A, Osterbur D, Ip I, Schneider L, et al. Fast Healthcare Interoperability Resources, Clinical Quality Language, and Systematized Nomenclature of Medicine-Clinical Terms in Representing Clinical Evidence Logic Statements for the Use of Imaging Procedures: Descriptive Study. JMIR Med Inform 2019 May 13;7(2):e13590 [FREE Full text] [doi: 10.2196/13590] [Medline: 31094359]
14. Berner ES, Kasiraman RK, Yu F, Ray MN, Houston TK. Data quality in the outpatient setting: impact on clinical decision support systems. AMIA Annu Symp Proc 2005:41-45 [FREE Full text] [Medline: 16778998]
15. Wong A, Amato MG, Seger DL, Rehr C, Wright A, Slight SP, et al. Prospective evaluation of medication-related clinical decision support over-rides in the intensive care unit. BMJ Qual Saf 2018 Sep;27(9):718-724. [doi: 10.1136/bmjqs-2017-007531] [Medline: 29440481]
16. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res 2018 May 29;20(5):e185 [FREE Full text] [doi: 10.2196/jmir.9134] [Medline: 29844010]

17. El-Sappagh S, Franda F, Ali F, Kwak K. SNOMED CT standard ontology based on the ontology for general medical science. BMC Med Inform Decis Mak 2018 Aug 31;18(1):76 [FREE Full text] [doi: 10.1186/s12911-018-0651-5] [Medline: 30170591]
18. Kalra D, Schulz S, Karlsson D, Vander SR, Cornet R, Rosenbeck GK, et al. Assessing SNOMED CT for Large Scale eHealth Deployments in the EU. Brussels: European Union; 2016. URL: http://www.assess-ct.eu [accessed 2020-10-01]
19. Williams R. A Christmas guide to clinical coding. BMJ 2018 Dec 13;363:k5209. [doi: 10.1136/bmj.k5209]
20. Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc 2006 Jun;81(6):741-748. [doi: 10.4065/81.6.741] [Medline: 16770974]

Abbreviations

ATC: Anatomical Therapeutic Chemical Classification
EBMeDS: Evidence-Based Medicine electronic Decision Support
ICD-10: International Classification of Diseases, Tenth Revision
ICPC: International Classification of Primary Care
LOINC: Logical Observation Identifiers Names and Codes
SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms

Edited by G Eysenbach; submitted 02.09.19; peer-reviewed by R Verheij, R Williams, E Ding; comments to author 21.08.20; revised version received 21.09.20; accepted 23.09.20; published 21.10.20.

Please cite as:
Delvaux N, Vaes B, Aertgeerts B, Van de Velde S, Vander Stichele R, Nyberg P, Vermandere M
Coding Systems for Clinical Decision Support: Theoretical and Real-World Comparative Analysis
JMIR Form Res 2020;4(10):e16094
URL: http://formative.jmir.org/2020/10/e16094/
doi:10.2196/16094
PMID:33084593

Original Paper

Digital Self-Management in Support of Patients Living With Chronic Pain: Feasibility Pilot Study

Katrine Bostrøm[1,2], MSc; Elin Børøsund[1], PhD; Cecilie Varsi[1], PhD; Hilde Eide[1,3], PhD; Elise Flakk Nordang[1], MSc; Karlein MG Schreurs[4], PhD; Lori B Waxenberg[5], PhD; Karen E Weiss[6], PhD; Eleshia J Morrison[7], PhD; Milada Cvancarova Småstuen[8], PhD; Audun Stubhaug[2,9,10], PhD; Lise Solberg Nes[1,2,7], PhD

[1]Department of Digital Health Research, Division of Medicine, Oslo University Hospital, Oslo, Norway
[2]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
[3]Science Centre Health and Technology, University of South-Eastern Norway, Drammen, Norway
[4]Department of Psychology, Health & Technology, University of Twente, Enschede, Netherlands
[5]Department of Clinical and Health Psychology, University of Florida, Gainesville, FL, United States
[6]Department of Anesthesiology and Pain Medicine, University of Washington School of Medicine, Seattle, WA, United States
[7]Department of Psychiatry and Psychology, College of Medicine and Science, Mayo Clinic, Rochester, MN, United States
[8]Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway
[9]Department of Pain Management and Research, Oslo University Hospital, Oslo, Norway
[10]Regional Advisory Unit on Pain, Oslo University Hospital, Oslo, Norway

Corresponding Author:
Lise Solberg Nes, PhD
Department of Digital Health Research
Division of Medicine
Oslo University Hospital
Pb 4950 Nydalen
Oslo, Oslo, N-0424
Norway
Phone: 47 91332341
Email: lise.solberg.nes@rr-research.no

Abstract

Background: Chronic pain can be complex and taxing to live with, and treatment and support require a multicomponent approach, which may not always be offered or available. Smartphones, tablets, and personal computers are already incorporated into patients' daily lives, and therefore, they can be used to communicate, educate, and support self-management. Although some web-based self-management interventions exist, research examining the evidence and effect of digital solutions supporting self-management for patients living with chronic pain is limited, findings are inconclusive, and new innovative ideas and solutions are needed.

Objective: This feasibility pilot study aimed to explore the system use, perceived usefulness, ease of use, and preliminary effects of EPIO, an app-based cognitive-behavioral pain self-management intervention program for patients living with chronic pain.

Methods: The EPIO intervention was delivered in a blended-care model containing (1) one face-to-face introduction session, (2) nine cognitive behavior–based pain self-management modules, delivered in an app-based format for smartphones or tablets, and (3) one follow-up phone call at 2 to 3 weeks after the introduction session. Patients living with chronic pain (N=50) completed pre-post outcome measures at baseline and 3 months after the introduction session, with registration of system use (ie, log data) until 6 months. The use, perceived usefulness, and ease of use of the EPIO program were examined through system use data, as well as a study-specific use/usability questionnaire and the System Usability Scale (SUS). Outcome measures to test feasibility of use and estimate preliminary effects included the Brief Pain Inventory, health-related quality of life (HRQoL) scale, Hospital Anxiety and Depression Scale, Self-Regulatory Fatigue scale, Pain Catastrophizing Scale, and Chronic Pain Acceptance Questionnaire.

Results: Participants (N=50) had a median age of 52 years (range 29-74 years) at inclusion and were mainly female (40/50, 80%). Thirty-one participants completed at least six of the nine modules within the 3-month study period (62% completion rate). Forty-five participants completed outcome measures at 3 months, and the EPIO program was rated as useful (ie, "totally agree"

or “agree”; 39/45, 87%) and easy to use (42/45, 93%), and as having easily understandable exercises (44/45, 98%). The average overall system usability (SUS) score was 85.7, indicating grade A and excellent system usability. Preliminary psychosocial outcome measure estimates showed primarily nonsignificant pre-post intervention improvements at 3 months, but with significant positive effects related to some aspects of HRQoL (bodily pain, *P*=.02 and change, *P*=.049).

Conclusions: Digital self-management intervention programs may be of use and support for patients living with chronic pain. In this feasibility study, EPIO showed an acceptable program completion rate and was rated as useful and easy to use, with excellent user satisfaction. Program optimization and efficacy testing in a large-scale randomized controlled trial are warranted and in progress.

Trial Registration: ClinicalTrials.gov NCT03705104; https://clinicaltrials.gov/ct2/show/NCT03705104

(JMIR Form Res 2020;4(10):e23893) doi:10.2196/23893

KEYWORDS

chronic pain; feasibility; acceptability; self-management; eHealth; digital; cognitive-behavioral pain; usability; user centered

Introduction

Background

Chronic pain is complex and taxing to live with, and how patients perceive and relate to pain is based on an interplay of biomedical, psychosocial, behavioral, and cultural factors [1]. Given this intricacy, chronic pain is optimally managed by treatments that address not only biological factors [2,3], but also psychological and social influences and consequences. The complexity, demands, and challenges of living with chronic pain may lead to a draining of capacity to self-regulate [4-6]. Helping patients to build or strengthen their self-regulatory capacity and support motivation to engage in pain self-management strategies can, therefore, be important [7]. International clinical guidelines also recommend the inclusion of self-management interventions in routine treatment for patients living with chronic pain [2].

Psychosocial interventions based on cognitive behavioral therapy (CBT) [8] and/or acceptance and commitment therapy (ACT) [9], aiming to support coping and self-management for patients living with chronic pain [10,11], have been shown to be associated with improved quality of life, pain acceptance, functioning, and self-efficacy, as well as reduced pain, anxiety, and depressive symptomatology [12-14]. Unfortunately, such individual or group in-person psychosocial interventions are not always an option for patients living with chronic pain [15]. Possible barriers include lack of accessibility of services, personal preferences, the medical condition itself, lack of insurance coverage, and geographical distance [16,17]. Given the limited availability and options of in-person psychosocial interventions for patients living with chronic pain, new and innovative ideas and solutions are needed [18].

Digital solutions in the form of eHealth solutions, defined as the use of digital communication-based technology to provide health care and support self-management of health conditions [19], may provide innovative options for patients living with chronic pain [20]. Patients with chronic pain have also reported being interested in eHealth interventions in support of self-management [7,21], and existing eHealth interventions for self-management of chronic pain have shown promise in terms of the potential to address unmet needs, support psychological well-being, strengthen self-efficacy, and increase flexibility [18,22]. However, findings and indications of efficacy for such pain-related interventions are still limited and mixed.

Most patient-oriented apps for people living with chronic pain only provide information about pain or about the illness, including ways to check symptoms and track medication use [23]. Few eHealth pain-related apps provide information about coping and self-management strategies [24,25], and even though some web-based CBT or ACT-based interventions have been tested in support of people living with chronic pain, findings are still inconclusive and interventions need further testing, also in app format [13,22,26-28]. Systematic reviews have concluded that eHealth interventions are more likely to be successful if developed with a user-centered focus, increasing the likelihood of matching the user’s needs and requirements [29,30]. However, systematic literature reviews examining the development and use of pain-related eHealth apps indicate that very few of these programs are developed with the involvement of health-care professionals and actual end users (ie, patients with chronic pain), and only a few existing pain-related apps appear to be based on a theoretical and evidence-based rationale [31,32]. These aspects emerge as major limitations of eHealth pain self-management interventions so far. In addition, few existing studies report system use and level of engagement (ie, app activity), and/or satisfaction/usability with eHealth pain self-management programs to date [33]. Given these challenges identified by existing scientific literature, new extensive research and innovative solutions are required to show the feasibility, usefulness/usability, and effectiveness of eHealth interventions supporting self-management for patients living with chronic pain [20,34].

In response to existing research and recommendations, the current research team has examined users’ (ie, patients living with chronic pain) [21] and health care providers’ [35] inputs related to needs and requirements for a potential eHealth pain self-management intervention [21,35]. Incorporating findings [21,35] and combining these with existing clinical and research evidence for the effectiveness of CBT/ACT-type interventions [36-39], the team subsequently designed and developed a cognitive-behavioral pain self-management eHealth intervention called *EPIO* (inspired by the Greek goddess for the soothing of pain, Epione), aiming to support patients living with chronic pain (ie, chronic pain in general, not pain/pain condition specific) [21,26,35]. This study builds on this research line.

Objectives

To enable the effective evaluation of complex interventions, the Medical Research Council recommends initial intervention testing and refinement to ensure intervention feasibility [40]. This feasibility pilot study therefore aimed to assess system use (ie, user app activity), perceived usefulness, and ease of use of the EPIO intervention program in order to identify needs for adjustments and to enable optimization in preparation for a future randomized controlled trial (RCT). The current feasibility pilot also aimed to explore preliminary efficacy findings (ie, pain interference, health-related quality of life [HRQoL], anxiety and depression, self-regulatory fatigue, pain catastrophizing, and pain acceptance) using unadjusted exploratory pre-post intervention analyses.

Methods

Description of the EPIO Intervention Program

The *EPIO* intervention program was designed and developed by a collaborative research team, consisting of scientists, health care providers, eHealth experts, content and system developers, and end-user representatives (ie, patients living with chronic pain) [26]. Content development [26] was based on well-known evidence-based aspects from CBT, with some integrated aspects of ACT, focusing on self-management and coping for patients living with chronic pain in general [7,11,15,30,41,42]. Focusing on chronic pain in general, the EPIO intervention program is so far not developed to be pain type/pain condition specific. The EPIO program contains nine modules designed with several interconnected parts of information and education (eg, pain physiology, coping strategies, thought challenges, and the importance of activity balance) and a variety of self-management–based exercises for patients living with chronic pain (eg, diaphragmatic breathing, graded behavioral activation, mindfulness, and progressive muscle relaxation) [26,41,43]. The nine modules in the EPIO program include the following topics: (1) information about pain, (2) balance, (3) thoughts and feelings, (4) stress and coping, (5) what is important to me (ie, values), (6) behaviors and lifestyle, (7) communication, relations, and social support, (8) coping during difficult times, and (9) summary and the road ahead [26].

To encourage program content practice, each module has to be open for 3 days (ie, practicing mode) before the next module will open. To provide structure and to allow individualization, the first five EPIO modules are sequential, while the order of modules 6 to 8 can be chosen. In addition, participants can create their own favorite list by highlighting exercises and can receive reminders according to their own needs. Participants can also choose between reading and listening to the program at any time. To ensure availability, the program can also be used while participants are offline. Details for the design, development, and content of EPIO are presented elsewhere [26]. The EPIO intervention is delivered in a blended-care model containing the following: (1) one face-to-face introduction session; (2) nine primarily CBT-based pain self-management modules [26], delivered in an app-based format for smartphones or tablets; and (3) one follow-up phone call conducted at 2 to 3 weeks after the introduction session.

Study Design

A pre-post intervention study without a control group was employed in this study, with all participants receiving the EPIO intervention. Outcome measures to test feasibility of use and derive estimates of preliminary efficacy were collected at baseline and at 3 months after the introduction session. In addition, data of system use (ie, log data) were collected for 6 months, and extracted at 3 and 6 months after the introduction session. Feasibility conceptualization was guided by Bowen et al [44] exploring (1) *acceptability* (to what extent is the EPIO program judged as suitable, satisfying, or attractive to program recipients); (2) *demand* (to what extent is EPIO likely to be used, ie, exploration of the actual use of the program); and (3) *limited efficacy testing* (does the EPIO program show promise of being successful with the intended population?).

Participants and Recruitment

Information about the study was communicated through the research project website [45], through the initiating institution (Oslo University Hospital), and through collaborating partners, including local health care services and primary care practices. Study and recruitment information was also advertised through social media channels and through patient organizations' web pages. The inclusion criteria were as follows: age ≥18 years, living with chronic pain in general (ie, not pain/pain condition specific), pain duration ≥3 months (self-reported), access to a smartphone or tablet, being able to understand oral and written Norwegian, and being able to attend an in-person introduction session. The exclusion criteria were as follows: having untreated severe mental illness, migraine, or cancer-related pain (all self-reported). Participants were recruited between January and May 2019.

Study Procedure

The study was approved by the Regional Committee for Medical and Health Research Ethics (REK 2018/8911) and the Oslo University Hospital Institutional Review Board equivalent function (PVO 2017/6697). All participants provided written informed consent. Participants attended one in-person introduction session where they were introduced to the self-management program in the introduction session and received help downloading the EPIO app from the App Store or Google Play Store, with instructions on how to get started. They also received a follow-up phone call from the study staff 2 to 3 weeks after the introduction session to see how things were going and whether there were any questions. If the participants had any additional questions or feedback (eg, technical issues) related to the EPIO program during the study, they could contact the study staff by phone. System use was logged for 6 months through a secure server at Services for Sensitive Data (TSD; University of Oslo). Outcome measures were completed through a secure TSD server online at baseline (ie, before the introduction session) and at 3 months after the introduction session (ie, at a 3-month follow-up). Program completers were defined as participants completing at least six out of the nine modules (67%) of the EPIO program in the 3-month study period [46].

System Use

For study purposes, participants were encouraged to try to complete all nine app-based modules within 3 months, but they could continue to use the program for as long as they wanted. System use (ie, user app activity) and program progress were automatically logged for 6 months through a secure server (ie, TSD) using an encrypted connection. To explore the extent to which the EPIO program was used by participants (ie, *demand*) [44], system use log data (ie, app activity) were extracted at 3 and 6 months after the introduction session.

Perceived Usefulness and Ease of Use

To explore the extent to which the EPIO program was judged as suitable, satisfying, or attractive to intervention recipients (ie, *acceptability*) [44], participants completed a six-item study-specific questionnaire (Multimedia Appendix 1), as well as the System Usability Scale (SUS) [47] at the 3-month follow-up. The study-specific questionnaire was based on previous experience with developing eHealth apps in the research team [26,48] and was guided by the Technology Acceptance Model (TAM) [49]. The first three items in the questionnaire, inspired by Davis [50], measure participant program perception as follows: (1) the program was easy to use, (2) the exercises were easy to understand, and (3) the program was useful. Response options range from *totally agree* to *totally disagree*. The remaining items in the study-specific questionnaire are open-ended questions designed to gather information related to participants' perceived usefulness and ease of use as follows: (4) What did you like the best? (5) What did you like the least? and (6) What are your suggestions for improvement? The SUS [47] is a 10-item questionnaire with five response options ranging from *strongly disagree* to *strongly agree*. SUS scoring yields a single number representing a composite measure of the overall usability of the system being studied. Scores are to be summarized and multiplied by 2.5, leading to a value range of 0 to 100, and 68 is considered the average score. A score above 80.3 can be interpreted as grade A (ie, the top 10% of scores), which equals excellent system usability [51].

Preliminary Effects: Outcome Measures

Outcome measures were collected at baseline (ie, before the introduction session) and at the 3-month follow-up. At baseline, patients also completed a study-specific demographic and disease-related measure. To test feasibility of use, explore preliminary effects, and assess whether the EPIO program showed promise of being successful in the intended patient population (ie, *limited efficacy testing*), participants completed several psychosocial outcome measures.

Pain interference was measured with seven items from the short form of the Brief Pain Inventory (BPI) [52], a measure of the impact of pain on daily function. The BPI has acceptable internal consistency and reliability and has been validated in a Norwegian chronic pain population sample [53]. The score range of BPI is 0 to 10, with higher scores indicating higher pain interference.

HRQoL was measured with the noncommercial SF-36 Short Form Health Survey (RAND-36) [54,55], a 36-item measure of physical, role, emotional, cognitive, and social function, as well as physical, general, and global health. The RAND-36 has acceptable internal consistency and reliability [54] and has been validated in a Norwegian population sample with chronic pain [55]. The score range of RAND-36 is 0 to 100, with higher subscale scores indicating better HRQoL.

Anxiety and depressive symptoms were measured with the Hospital Anxiety and Depression Scale (HADS) [56], a 14-item measure of anxiety and depressive symptomatology, validated as a unidimensional measure of psychological distress. The HADS has acceptable internal consistency and reliability [56]. The score range of HADS is 0 to 21 for both scales, with higher scores indicating a higher presence of anxiety or depression.

Self-regulatory fatigue was measured with the Self-Regulatory Fatigue 18 (SRF-18) [57], an 18-item scale measuring self-regulatory capacity with cognitive, emotional, and behavioral components. The SRF-18 has acceptable internal consistency and reliability [57]. The score range of SRF-18 is 18 to 90, with higher scores indicating higher self-regulatory fatigue.

Pain catastrophizing was measured with the Pain Catastrophizing Scale (PCS) [58,59], a 13-item scale measuring catastrophic thinking and maladaptive responses to pain. Three subscales measure helplessness, magnification, and rumination. The PCS has acceptable internal consistency and reliability, and has been validated in a Norwegian population sample with chronic pain [59]. The score range of the PCS is 0 to 52, with higher scores indicating higher catastrophic thoughts and feelings about pain.

Pain acceptance was measured with the short form of the Chronic Pain Acceptance Questionnaire (CPAQ-8) [60,61], an eight-item scale measuring pain acceptance. The CPAQ-8 has acceptable internal consistency and reliability [61], and has been validated in a Norwegian population sample with chronic pain [60]. The score range of the CPAQ-8 is 0 to 24, with higher scores indicating a higher acceptance of pain.

Qualitative Analyses

Qualitative data from the open-ended questions in the study-specific questionnaire were analyzed using an Excel spreadsheet, according to a thematic analysis process (ie, coding reliability) as described by Braun and Clark [62]. The first author (KB) performed the analyses of the data, in collaboration with a coauthor (EB). The data were grouped as domains, directly guided by the study questions (ie, what did you like the best, what did you like the least, and suggestions for improvement), before categories were derived.

Statistical Analyses

Data were analyzed using the Statistical Program for Social Sciences (SPSS), version 25 (IBM Corp). Data on baseline characteristics, system log data, and usefulness/ease of use data are presented as medians and ranges for continuous variables and as proportions with percentages for categorical variables. Paired samples *t* tests were used to assess possible pre-post intervention changes. To explore potential group differences in outcome measures, demographics, and program progress,

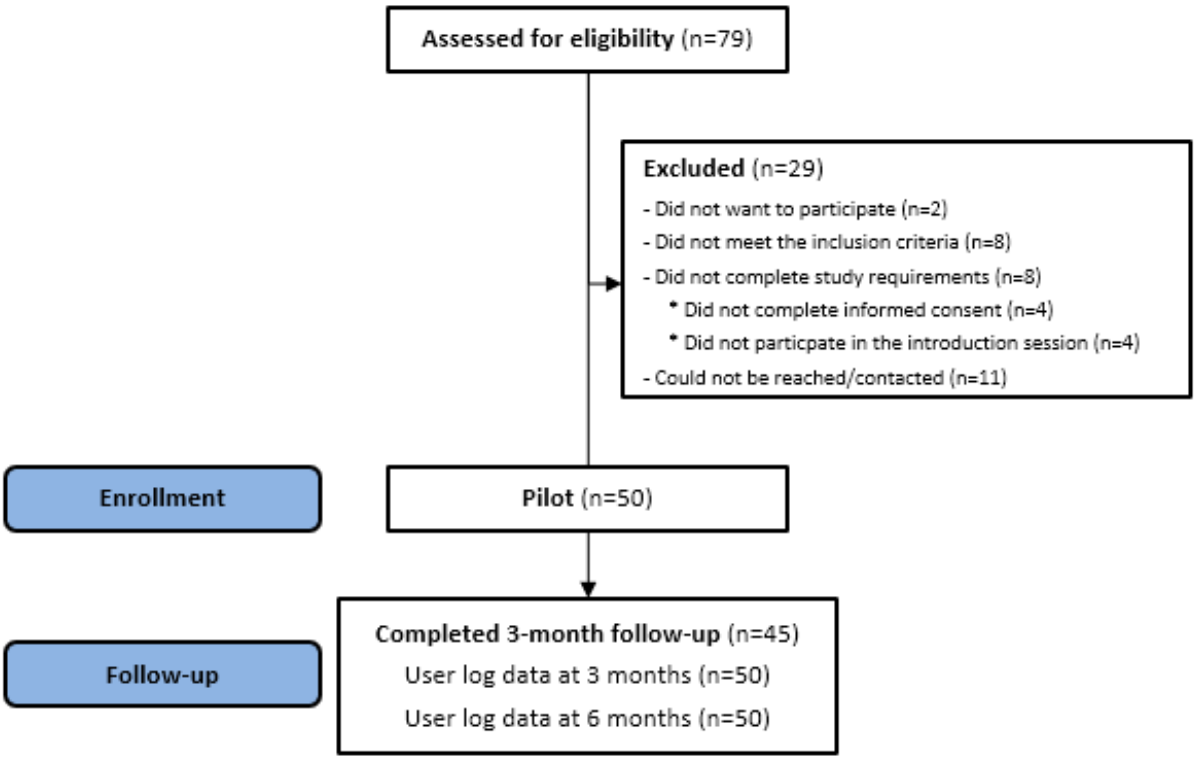
univariate linear regression analyses were conducted. All statistical tests were two-sided. *P* values <.05 were considered statistically significant. As this was a feasibility pilot study, results were considered exploratory and no correction for multiple testing was performed [63].

Results

Recruitment, Participant Flow, and Sample Description

Between January and May 2019, 79 patients living with chronic pain were referred for potential study participation. Of these, 29 were excluded (ie, did not meet the inclusion criteria, did not complete study requirements, could not be reached, or declined to participate). Fifty patients with chronic pain were included in the study and received the EPIO intervention. Forty-five of the initial participants completed the 3-month outcome measures. Of the five participants who did not complete the 3-month outcome measures, two described a worsened health condition as the reason. No other reasons were provided. Apart from age (ie, noncompleters were significantly older than completers, 62.2 years versus 51.3 years, *P*=.02), there were no differences between noncompleters and completers with regard to baseline variables. Figure 1 provides a summary of the recruitment and participant flow.

Figure 1. Recruitment and participant flow.



Participants (N=50) were primarily Caucasian (48/50, 96%), had a median age of 52 years (range 29-74 years) at inclusion, and were mainly female (40/50, 80%). Forty-one participants provided self-reported details related to their pain conditions, including pain related to unspecific musculoskeletal pain (eg, back and neck pain [16/41, 39%], unspecified disc disorder [8/41, 19%], osteoarthritis [7/41, 17%], fibromyalgia [7/41, 17%], neuropathy [7/41, 17%], complex regional pain syndrome [4/41, 10%], injuries [5/41, 12%], and surgeries [3/41, 7%]), with more than half of the participants reporting more than one diagnosis (23/41, 56%). The majority (37/50, 74%) of the participants reported having lived with pain for 5 years or longer. Table 1 presents the baseline demographics and illness characteristics of the participants.

Table 1. Baseline demographics and illness characteristics (N=50).

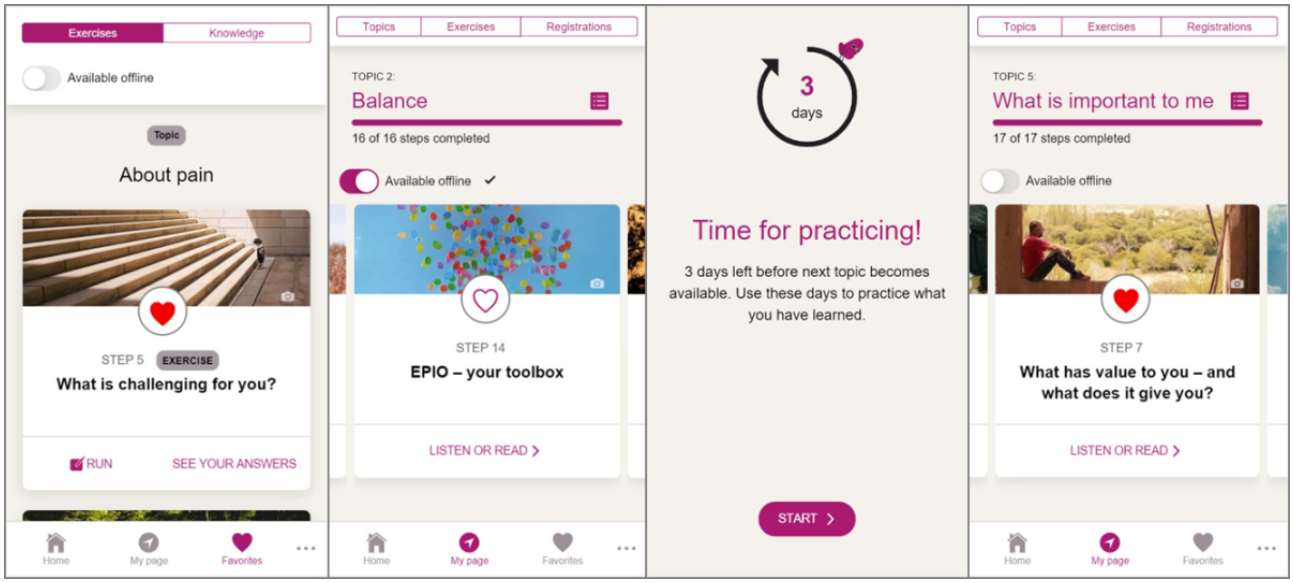
Variable	Value, n (%)
Gender	
Female	40 (80)
Male	10 (20)
Marital status	
Married/cohabitating	29 (58)
Single/divorced	21 (42)
Education	
Elementary/high school	17 (34)
University/college <4 years	21 (42)
University/college ≥4 years	12 (24)
Employment	
Full-time/part-time work	14 (28)
Sick leave/disability benefits	29 (58)
Retired/others	7 (14)
Years living with pain	
1-3 years	10 (20)
3-5 years	3 (6)
5-10 years	13 (26)
>10 years	24 (48)
Health services usage[a]	
General practitioner	47 (94)
Physiotherapy/physical therapy	40 (80)
Psychology	16 (32)
Pain physician/pain specialist services	6 (12)
Pain clinic	16 (32)
Occupational therapy	3 (6)
Rehabilitation	16 (32)
Healthy life centers	3 (6)
Educational courses	9 (18)
Other	11 (22)

[a]Participants could report having received several types of health services during the course of their illness.

System Use

Thirty-one participants completed at least six of the nine EPIO program modules within the 3-month study period, yielding a 62% intervention completion rate at 3 months. Fourteen (28%) participants completed all nine modules within 3 months. Noncompleters (ie, completing less than six modules in the 3-month study period) completed an average of two modules in 3 months. Of the EPIO intervention program exercises, the top three exercises repeated most frequently during the 3-month study period were as follows: “What is challenging to you,” “Choice of fun activities,” and ”What kind of thoughts do you use.” For example screenshots of the EPIO program, refer to Figure 2.

Figure 2. Example screenshots from the EPIO intervention program. From the left: (1) Exercise example about daily challenges; (2) Module about activity pacing; (3) Practicing mode; and (4) Module about values.



While the pre-post intervention study period was 3 months for practical purposes, system use was monitored for 6 months. At the 6-month follow-up, 32 (64%) participants had completed at least six out of the nine program modules. Nineteen (38%) participants had completed all nine modules at the 6-month follow-up.

Perceived Usefulness and Ease of Use

At the 3-month follow-up, 45 participants (90%) completed measures related to their satisfaction with the EPIO program. In the study-specific use/usability questionnaire, the EPIO program was rated as useful (ie, *totally agree* or *agree*; 39/45, 87%) and easy to use (42/45, 93%), and as having easily understandable exercises (44/45, 98%). Main findings from the open-ended questions showed that a majority of the participants reported appreciating the exercises (eg, the relaxation and diaphragmatic breathing exercises), the combination of exercises and educational information, the easy access, and the functionality of being able to choose between reading and listening or being able to do both. The mean system usability (ie, SUS) score was 85.7 (SD 12.9), indicating grade A, which equals excellent (ie, score >80.3) system usability. Even though few men participated in this study, there were some indications of gender differences, with more women (85%) than men (50%) rating the program as useful. Additionally, there were some differences related to education, with higher educated participants (ie, >4 years of university) having significantly higher SUS scores than lower educated participants (ie, elementary/high school, P=.04).

Preliminary Effects: Pre-Post Interventions Results

Preliminary pre-post intervention findings at the 3-month postintroduction session did not reach statistical significance for the majority of the psychosocial outcome measures (details are provided in Table 2). HRQoL findings indicated statistically significant improvements from baseline to postintervention for “bodily pain” (mean difference [MD] 5.1; P=.02) and the single item “change” (ie, perceived change in health) (MD 5.6; P=.049), and there was a trend toward significant improvements for the Role Physical scale, but the result was not statistically significant (MD 10.0; P=.07). There was a high degree of heterogeneity in the data, reflected in large values of variance and subsequently broad CIs for the point estimates in a number of subscales (eg, HRQoL Role Emotional with a 10-point positive score change, but with CI −3.3 to 24.0 and consequently no statistical significance, P=.13). There was also a trend toward significant reduced pain catastrophizing, but the result was not statistically significant (MD 1.8; P=.06). Scores related to anxiety, depression, and self-regulatory fatigue remained stable. Moreover, the results after 3 months indicated statistically significant lower pain acceptance (“Willingness to accept” subscale, MD 0.9; P=.03; total score, MD −1.4; P=.02).

Table 2. Preliminary effects: pre-post intervention changes in psychosocial outcomes (n=45).

Psychosocial outcomes	Baseline, mean (SD)	3-month follow-up, mean (SD)	Mean difference (95% CI)	*P* value
Pain interference (BPI[a])[b]	4.5 (2.1)	4.8 (2.1)	0.3 (−0.3 to 0.8)	.31
HRQoL[c] (RAND-36[d])				
Physical function	60.9 (24.6)	61.3 (25.4)	0.4 (−3.1 to 3.9)	.80
Role physical	17.2 (31.5)	27.2 (36.5)	10.0 (−0.9 to 20.9)	.07
Bodily pain	35.3 (13.4)	40.4 (15.9)	5.1 (0.7 to 9.5)	.02
General health	46.3 (19.6)	48.9 (19.7)	2.6 (−1.2 to 6.3)	.18
Vitality	31.9 (21.0)	34.1 (20.8)	2.2 (−2.4 to 6.8)	.34
Social function	50.3 (24.9)	52.8 (24.1)	2.5 (−2.4 to 7.4)	.31
Role emotional	56.3 (45.4)	66.7 (40.8)	10.4 (−3.3 to 24.0)	.13
Mental health	63.7 (18.1)	65.1 (17.9)	1.3 (−1.2 to 3.8)	.29
Change	49.4 (25.3)	55.0 (23.0)	5.6 (0.0 to 11.1)	.049
Anxiety (HADS-A[e])	7.9 (3.6)	8.0 (4.0)	0.0 (−0.8 to 0.9)	.96
Depression (HADS-D[f])	5.8 (3.2)	5.8 (3.9)	0.1 (−0.6 to 0.6)	>.99
Self-regulatory fatigue (SRF-18[g])	53.7 (9.2)	53.3 (9.2)	0.4 (−2.5 to 1.6)	.66
Pain catastrophizing (PCS[h])				
Rumination	7.4 (3.9)	6.9 (4.0)	0.6 (−1.3 to 0.2)	.12
Magnification	3.3 (2.4)	3.0 (2.2)	0.3 (−0.8 to 0.1)	.14
Helplessness	8.4 (4.8)	7.6 (5.2)	0.9 (−1.9 to 0.2)	.11
Total score	19.2 (9.9)	17.4 (10.4)	1.8 (−3.7 to 0.1)	.06
Chronic pain acceptance (CPAQ[i])				
Willingness	14.0 (2.8)	13.1 (3.1)	0.9 (−1.7 to −0.1)	.03
Activity engagement	14.3 (3.3)	13.8 (3.3)	0.5 (−1.4 to 0.4)	.23
Total score	28.3 (5.2)	26.9 (5.3)	1.4 (−2.7 to −0.2)	.02

[a]BPI: Brief Pain Inventory.

[b]Subscale of the Brief Pain Inventory (score range 0-10; a higher score indicates higher interference in life).

[c]HRQoL: health-related quality of life.

[d]RAND-36: RAND 36-Item scale (score range 0-100; a higher score indicates higher emotional well-being).

[e]HADS-A: Hospital Anxiety and Depression Scale-Anxiety subscale (score range 0-21; a higher score indicates a higher degree of anxiety).

[f]HADS-D: Hospital Anxiety and Depression Scale-Depression subscale (score range 0-21; a higher score indicates a higher degree of depression).

[g]SRF-18: Self-regulatory Fatigue 18 scale (score range 18-90; a higher score indicates higher self-regulatory fatigue).

[h]PCS: Pain Catastrophizing Scale (score range 0-52; a higher score indicates higher catastrophizing).

[i]CPAQ: Chronic Pain Acceptance Questionnaire (score range 0-52; a higher score indicates a higher acceptance of pain).

Program Completion and Relations to Psychosocial Outcomes

The participants who completed the EPIO program (ie, completed six or more modules) within the 3-month study period had higher mean pain interference and lower emotional well-being scores at baseline compared with those who did not complete the program (ie, completed five or less modules). However, these differences were not statistically significant (both $P=.28$).

Preparation for Optimization and a Randomized Controlled Trial

In order to optimize the EPIO intervention and prepare for efficacy testing in an RCT, system use, perceived usefulness, and ease of use findings from this study were employed to prioritize and address the needs for further adjustments to the EPIO intervention program after the feasibility pilot. For example, as the 3-day delay between modules (ie, practicing mode) received mixed feedback, with some preferring to move forward more rapidly and others liking the practicing mode, modules 1 and 2 were set to open simultaneously as the introduction session and module 1 had overlapping themes.

Additionally, based on system log data showing increased program activity right before and right after the 2 to 3 week follow-up phone call, a decision was made to add a second follow-up phone call (eg, at 6-7 weeks) for the future RCT.

Discussion

Feasibility of eHealth in Chronic Pain

Feasibility studies play an important role in the planning of RCTs to examine novel interventions or to examine a combination of existing interventions in new patient populations or recruitment settings [44]. Patients with chronic pain have reported being interested in eHealth interventions in support of self-management [7,21]. Given the early stage of evidence-based eHealth interventions, however, the need for more studies reporting on the feasibility, usability, and efficacy of eHealth intervention programs for patients living with chronic pain is evident [20,34]. This feasibility pilot study therefore examined the system use, perceived usefulness, ease of use, and preliminary effects of EPIO, an app-based cognitive-behavioral pain self-management intervention for patients living with chronic pain (ie, chronic pain in general, not pain/pain condition specific).

Principal Findings

In the 3-month study-period, 62% (31/50) of participants completed at least six out of nine modules of the EPIO program. The participants rated the program as useful (39/45, 87%) and easy to use (42/45, 93%), and mentioned the presence of easily understandable exercises (44/45, 98%). System usability was rated as excellent, and although mainly nonsignificant (details are provided in Table 2), preliminary psychosocial outcome measures indicated some positive impact related to HRQoL. The repeated use of multiple self-management exercises in the EPIO program also suggests that exercise variety may be particularly of interest and may support program use and engagement for patients living with chronic pain. As suggested by scientific literature reviews examining status and limitations of eHealth interventions [20,32,34], the EPIO intervention program was developed based on existing evidence, in close collaboration among scientists, user representatives, and health care providers [26], which may have contributed to the positive feasibility and acceptability, including system use, perceived usefulness, and ease of use findings, noted in this study.

Participants' engagement is a precondition for the effectiveness of self-management interventions [64]. In this feasibility pilot, participants were therefore encouraged to spend as much time as possible becoming familiar with the EPIO program and to practice the content and variety of exercises as much as possible. The 62% completion rate (ie, completing at least six modules) during the 3-month study period is likely lower than the completion rates for equivalent in-person interventions. However, adherence/completion rates have emerged as a challenge for eHealth interventions, sometimes being as low as 20% to 40%, and the 62% completion rate in this study can therefore be considered acceptable.

System log data examinations also revealed slightly higher program completion rates at the 6-month follow-up than at the 3-month follow-up. Research indicates that the complexity, demands, and challenges of living with chronic pain may lead to a draining of the capacity to self-regulate cognitive, emotional, and behavioral activities, sometimes particularly related to executing functioning [4-6]. If patients living with chronic pain struggle with the many important decisions and the behavioral changes often required for successful self-management, finding new ways to support patients with chronic pain to maintain engagement and adhere to self-management programs may be important [65]. Given the findings from this study, it is possible that patients living with chronic pain could benefit from having more time (ie, >3 months) to process the program information than initially estimated by this research team. Participants in this study were however given access to the EPIO program for as long as they wanted. Therefore, despite the limited increase in use and completion rates from 3 to 6 months, the apparent continued use of the EPIO program in the poststudy period (ie, >3 months) may indicate that participants took their time engaging in the EPIO intervention program and found the program useful after study completion.

Program completers in this study reported higher pain interference and lower emotional well-being at baseline compared with noncompleters. This could potentially suggest that higher levels of pain interference and lower emotional well-being may be positively associated with program interest, motivation for change, engagement, and completion. Such indications could be of great interest for the development and use of future eHealth interventions. However, in this feasibility pilot study, the associations among program completion, progress, and psychosocial outcomes were nonsignificant (P=.28). Conclusions cannot be made, and this issue needs to be further explored by future research in RCTs with larger study populations.

Despite some indications that older participants may have been more likely to be noncompleters in this study, only five individuals were study noncompleters by 3 months, and conclusions cannot be made. There are however some suggestions that high age could be associated with lower eHealth use, with younger people expressing more willingness and interest in using eHealth and older adults expressing worry about losing personal contact with their physicians if they start to use eHealth [66]. Research nevertheless also suggests that older adults seem to adhere to eHealth technology longer than younger people after starting to use such technology [66]. The link between age and eHealth use is not clear, as several studies have failed to show any difference in eHealth use based on age, and more research in this area is needed [66].

The outcomes of eHealth self-management interventions seem to depend on patients' motivation and adherence [7,25,64,67]. The adherence issues related to eHealth interventions are of great concern and need to be addressed in order for end users to achieve the intended intervention benefit. There are some indications that eHealth interventions may yield better adherence and subsequent effects when combined with face-to-face/in-person support [64]. The EPIO intervention was delivered in a blended-care model, which aimed to increase the motivation and likelihood of acceptance and completion of the

self-management intervention. It is therefore possible that the introduction session and follow-up phone call helped increase acceptability and program engagement in this study. One way to improve eHealth intervention adherence and completion rates could therefore be to increase the level of contact between health care/research study personnel and participants throughout the intervention [65], which is also one reason the research team decided to change from one to two follow-up phone calls for the planned future RCT.

As the primary aim of this study was to examine feasibility, including system use, usability, and ease of use, psychosocial outcome measures were primarily included to test feasibility of use. Potential pre-post effects were only examined as preliminary indications, and findings did not yield relevant results. Despite limited findings, the potential positive impact on HRQoL can be considered promising. Additionally, data variability was large, which may indicate that even though some participants may not have benefited greatly from the EPIO intervention program, others may have benefited greatly. The statistically significant (P=.02) finding showing reduced pain acceptance at the 3-month follow-up in this study is challenging to interpret. However, given that almost half (24/50, 48%) of the participants in the study reported having lived with chronic pain for more than 10 years, it may be overly optimistic to expect improvement in pain acceptance, particularly that the "willingness to accept" subscale would be impacted in such a short time. On the other hand, the trend toward reduced pain catastrophizing could be a positive indication. Future research should pay attention to these issues though, preferably also undertaking qualitative interviews to aid in intervention program use, usability, and effect interpretations.

Study Limitations and Strengths

This study has some limitations. First, the study was designed to assess the feasibility of a digital self-management intervention program in support of patients living with chronic pain; therefore, participants were not randomized, all participants received the intervention, and no definitive statements regarding the effectiveness of the intervention could be made. The study did however successfully establish feasibility and acceptability as intended, with acceptable system use and excellent perceived usefulness and ease of use. Second, the participants were recruited through social media and collaborating partners, and it may be assumed that the study population consisted of highly motivated people. This study therefore cannot conclude whether patients living with chronic pain would in general be interested in or benefit from such an intervention. The strong indications of feasibility in this study, with high acceptability/usability, are however promising. It should also be noted that as the EPIO program is designed to support patients living with chronic pain in general (ie, not pain condition specific), this study was not designed to examine differences in results based on the pain condition. The majority of the participants in this study were Caucasian, women, and those with higher education. People with higher education may be more likely to use health apps in general, while women may be more likely to use self-care or self-management apps and tend to exhibit the highest adherence to digital interventions [68]. Future studies should strive to incorporate user testing and recruitment strategies that may include a wider representation of potential end user groups (eg, gender, education, and ethnicity) in order to further test generalizability.

This study has several strengths. The EPIO intervention is built on clinical and research-based scientific evidence, and is designed and developed in collaboration with end users and related health care personnel, which is a definite strength. The extensive data gathering of 6-month logged system use, as well as usefulness and ease of use examinations are important study strengths. In addition, the inclusion of participants with a variety of pain-related diagnoses can be considered a strength. The goal of the EPIO intervention is to contribute to reducing the negative impact of chronic pain, no matter which type of pain patients experience, and as such, the current feasibility findings may be indicative for patients living with chronic pain in the general population.

Future Directions

This study established the feasibility of the digital EPIO pain self-management intervention. Suggestions for adjustments needed for optimizing and preparing a future RCT were made, and qualitative interviews for further data exploration were suggested. Additionally, the type of blended care delivery used in this study has the potential to enhance accessibility and actual use of psychosocial interventions and to enhance outreach for patients living with chronic pain. As adherence continuously appears to represent a major challenge for the success of eHealth interventions, future research should explore how aspects of design and development (eg, user involvement and prototype adherence testing) as well as delivery (eg, blended care, continuous follow-up, and social support features) could strengthen intervention adherence in future studies.

Given the mixed findings in the literature related to the utility of eHealth interventions for older adults [66], future research should also aim to incorporate ways to help older adults adopt eHealth interventions. This could potentially be done by including older adults in the design and development process of eHealth interventions [69], but perhaps even more importantly, providing proper training and education for participants on introducing such interventions [69,70]. A future EPIO intervention RCT should also explore this issue further. As system log data examinations revealed continued use and slightly higher program completion rates at the 6-month than 3-month follow-up, future research should examine the preferred and perhaps most likely to be effective period of intervention program use for patients living with chronic pain. Finally, rather than attempting to interpret efficacy outcomes from feasibility findings, efficacy must be evaluated in a future large-scale RCT with long-term follow-up. In the future, examining whether interventions, such as EPIO, have more or less potential for impact based on the type of pain/pain condition targeted is of importance.

Conclusions

This feasibility pilot study showed how digital self-management intervention programs, such as EPIO, a cognitive-behavioral eHealth pain self-management intervention, may be of use and support for patients living with chronic pain. EPIO program

completion rates were acceptable, program feasibility and acceptability were established, and the program was rated as useful and easy to use, with excellent user satisfaction. Intervention program optimization and efficacy testing in a large-scale RCT are warranted and in progress.

Acknowledgments

This study was supported by the Norwegian Research Council (grant number: 256574; principal investigator: Lise Solberg Nes). The authors would like to thank the patients who participated in the study. Without them, there would be no such research. The authors would also like to thank the leaders and health care providers at the learning and mastery centers at Vestre Viken Hospital Trust, the municipality care in Drammen, the Rheumatic Hospital at Lillehammer, and other participating providers for enabling recruitment and supporting this project. Finally, the authors would like to thank the research project team members, including the content development group and the design and software team at the Department of Digital Health Research at Oslo University Hospital, for their exceptional work throughout the project process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A study-specific questionnaire for perceived usefulness and ease of use.
[DOCX File , 34 KB - formative_v4i10e23893_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1062 KB - formative_v4i10e23893_app2.pdf]

References

1. Dansie EJ, Turk DC. Assessment of patients with chronic pain. Br J Anaesth 2013 Jul 24;111(1):19-25 [FREE Full text] [doi: 10.1093/bja/aet124] [Medline: 23794641]
2. Mills S, Torrance N, Smith BH. Identification and Management of Chronic Pain in Primary Care: a Review. Curr Psychiatry Rep 2016 Feb;18(2):22 [FREE Full text] [doi: 10.1007/s11920-015-0659-9] [Medline: 26820898]
3. Breivik H, Eisenberg E, O'Brien T, OPENMinds. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. BMC Public Health 2013 Dec 24;13(1):1229 [FREE Full text] [doi: 10.1186/1471-2458-13-1229] [Medline: 24365383]
4. Solberg Nes L, Roach AR, Segerstrom SC. Executive functions, self-regulation, and chronic pain: a review. Ann Behav Med 2009 Apr;37(2):173-183. [doi: 10.1007/s12160-009-9096-5] [Medline: 19357933]
5. Hamilton NA, Karoly P, Kitzman H. Self-Regulation and Chronic Pain:The Role of Emotion. Cognitive Therapy and Research 2004 Oct;28(5):559-576 [FREE Full text] [doi: 10.1023/b:cotr.0000045565.88145.76]
6. Solberg Nes L, Carlson CR, Crofford LJ, de Leeuw R, Segerstrom SC. Self-regulatory deficits in fibromyalgia and temporomandibular disorders. Pain 2010 Oct;151(1):37-44. [doi: 10.1016/j.pain.2010.05.009] [Medline: 20561749]
7. Huygens MW, Vermeulen J, Swinkels IC, Friele RD, van Schayck OC, de Witte LP. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Jul 08;16:232 [FREE Full text] [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]
8. Beck AT. Cognitive therapy and the emotional disorders. New York, USA: International Universities Press; 1976.
9. Hayes SC. Acceptance and Commitment Therapy, Relational Frame Theory, and the Third Wave of Behavioral and Cognitive Therapies - Republished Article. Behav Ther 2016 Nov;47(6):869-885. [doi: 10.1016/j.beth.2016.11.006] [Medline: 27993338]
10. Knoerl R, Lavoie Smith EM, Weisberg J. Chronic Pain and Cognitive Behavioral Therapy: An Integrative Review. West J Nurs Res 2016 May;38(5):596-628. [doi: 10.1177/0193945915615869] [Medline: 26604219]
11. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
12. Eccleston C, Crombez G. Advancing psychological therapies for chronic pain. F1000Res 2017;6:461 [FREE Full text] [doi: 10.12688/f1000research.10612.1] [Medline: 28413627]
13. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KM. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
14. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain:

randomized controlled trial. J Med Internet Res 2013 Jan 07;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
15. Kress H, Aldington D, Alon E, Coaccioli S, Collett B, Coluzzi F, et al. A holistic approach to chronic pain management that involves all stakeholders: change is needed. Curr Med Res Opin 2015;31(9):1743-1754. [doi: 10.1185/03007995.2015.1072088] [Medline: 26172982]
16. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
17. Jerant AF, von Friederichs-Fitzwater MM, Moore M. Patients' perceived barriers to active self-management of chronic conditions. Patient Educ Couns 2005 Jun;57(3):300-307. [doi: 10.1016/j.pec.2004.08.004] [Medline: 15893212]
18. Grace-Farfaglia P. Social Cognitive Theories and Electronic Health Design: Scoping Review. JMIR Hum Factors 2019 Jul 19;6(3):e11544 [FREE Full text] [doi: 10.2196/11544] [Medline: 31325290]
19. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]
20. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An Evaluation of the Effectiveness of the Modalities Used to Deliver Electronic Health Interventions for Chronic Pain: Systematic Review With Network Meta-Analysis. J Med Internet Res 2019 Jul 17;21(7):e11086 [FREE Full text] [doi: 10.2196/11086] [Medline: 31317869]
21. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' Needs and Requirements for eHealth Pain Management Interventions: Qualitative Study. J Med Internet Res 2019 Apr 01;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]
22. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: A systematic review. Internet Interv 2016 May;4:17-34 [FREE Full text] [doi: 10.1016/j.invent.2015.12.001] [Medline: 30135787]
23. Zhao P, Yoo I, Lancey R, Varghese E. Mobile applications for pain management: an app analysis for clinical usage. BMC Med Inform Decis Mak 2019 May 30;19(1):106 [FREE Full text] [doi: 10.1186/s12911-019-0827-7] [Medline: 31146739]
24. Martorella G, Gélinas C, Bérubé M, Boitor M, Fredericks S, LeMay S. The effect of tailored Web-based interventions on pain in adults: a systematic review protocol. Syst Rev 2016 Apr 12;5:59 [FREE Full text] [doi: 10.1186/s13643-016-0233-5] [Medline: 27072140]
25. Wilson M, Roll JM, Corbett C, Barbosa-Leiker C. Empowering Patients with Persistent Pain Using an Internet-based Self-Management Program. Pain Manag Nurs 2015 Aug;16(4):503-514. [doi: 10.1016/j.pmn.2014.09.009] [Medline: 26088940]
26. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KM, et al. A User-Centered Approach to an Evidence-Based Electronic Health Pain Management Intervention for People With Chronic Pain: Design and Development of EPIO. J Med Internet Res 2020 Jan 21;22(1):e15889 [FREE Full text] [doi: 10.2196/15889] [Medline: 31961331]
27. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. J Behav Med 2019 Apr 01;42(2):169-187. [doi: 10.1007/s10865-018-9984-x] [Medline: 30387008]
28. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of Mobile Apps in Pain Management: Systematic Review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
29. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Numans ME, Chavannes NH. From chronic disease management to person-centered eHealth; a review on the necessity for blended care. Clinical eHealth 2018 Mar;1(1):3-7. [doi: 10.1016/j.ceh.2018.01.001]
30. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
31. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the Quality of Mobile Apps for the Management of Pain: Systematic Search and Evaluation Using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
32. Enam A, Torres-Bonilla J, Eriksson H. Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
33. Lee J, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
34. Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A Systematic Review of Technology-assisted Self-Management Interventions for Chronic Pain. The Clinical Journal of Pain 2015;31(6):470-492. [doi: 10.1097/ajp.0000000000000185]
35. Heldal K, Varsi C, Ledel SI, Kristjansdottir O, Eide H, Mirkovic J. Building an e-health program for chronic pain management; Experiences and input from health care providers. Annals of Behavioral Medicine 2018;52(1):s98. [doi: 10.1093/abm/kay013]
36. Bruce B, Harrison TE, editors. Mayo Clinic Guide to Pain Relief. New York, USA: Mayo Clinic; 2014.

37. Veehof M, Oskam MJ, Schreurs KM, Bohlmeijer E. Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542. [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
38. Stanos S. Focused review of interdisciplinary pain rehabilitation programs for chronic pain management. Curr Pain Headache Rep 2012 Apr;16(2):147-152. [doi: 10.1007/s11916-012-0252-4] [Medline: 22427179]
39. Waxenberg LB. Gainesville, Florida: Department of Clinical and Health Psychology, College of Public Health and Health Professions, University of Florida; 2008.
40. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
41. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KM. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. Cogn Behav Ther 2016;45(1):5-31. [doi: 10.1080/16506073.2015.1098724] [Medline: 26818413]
42. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
43. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014 Feb 26(2):CD010152 [FREE Full text] [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]
44. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
45. EPIO. URL: www.epio.no [accessed 2020-10-06]
46. Børøsund E, Ehlers SL, Varsi C, Clark MM, Andrykowski MA, Cvancarova M, et al. Results from a randomized controlled trial testing StressProffen; an application-based stress-management intervention for cancer survivors. Cancer Med 2020 Jun;9(11):3775-3785 [FREE Full text] [doi: 10.1002/cam4.3000] [Medline: 32243717]
47. Brook J. SUS: A quick and dirty usability scale. In: Jordan PW, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London, UK: Taylor & Francis; 1996:94.
48. Børøsund E, Varsi C, Clark MM, Ehlers SL, Andrykowski MA, Sleveland HR, et al. Pilot testing an app-based stress management intervention for cancer survivors. Transl Behav Med 2020 Aug 07;10(3):770-780 [FREE Full text] [doi: 10.1093/tbm/ibz062] [Medline: 31330023]
49. Lee Y, Kozar KA, Larsen KR. The Technology Acceptance Model: Past, Present, and Future. CAIS 2003;12(50):751-781. [doi: 10.17705/1CAIS.01250]
50. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
51. Sauro J, Lewis J. When designing usability questionnaires, does it hurt to be positive? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 2011; Vancover, BC, Canada p. 2215-2224. [doi: 10.1145/1978942.1979266]
52. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. Clin J Pain 2004;20(5):309-318. [doi: 10.1097/00002508-200409000-00005] [Medline: 15322437]
53. Klepstad P, Loge JH, Borchgrevink PC, Mendoza TR, Cleeland CS, Kaasa S. The Norwegian brief pain inventory questionnaire: translation and validation in cancer pain patients. J Pain Symptom Manage 2002 Nov;24(5):517-525 [FREE Full text] [doi: 10.1016/s0885-3924(02)00526-2] [Medline: 12547051]
54. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
55. Loge JH, Kaasa S, Hjermstad MJ, Kvien TK. Translation and performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. J Clin Epidemiol 1998 Nov;51(11):1069-1076. [doi: 10.1016/s0895-4356(98)00098-5] [Medline: 9817124]
56. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
57. Nes LS, Ehlers SL, Whipple MO, Vincent A. Self-regulatory fatigue in chronic multisymptom illnesses: scale development, fatigue, and self-control. J Pain Res 2013;6:181-188 [FREE Full text] [doi: 10.2147/JPR.S40014] [Medline: 23526193]
58. Sullivan MJ, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
59. Fernandes L, Storheim K, Lochting I, Grotle M. Cross-cultural adaptation and validation of the Norwegian pain catastrophizing scale in patients with low back pain. BMC Musculoskelet Disord 2012 Jun 22;13(1):111 [FREE Full text] [doi: 10.1186/1471-2474-13-111] [Medline: 22726668]
60. Eide H, Leren L, Sørebø Ø. The Norwegian versions of the Chronic Pain Acceptance Questionnaire CPAQ-20 and CPAQ-8 - validation and reliability studies. Disabil Rehabil 2017 Jul 03;39(14):1441-1448. [doi: 10.1080/09638288.2016.1198427] [Medline: 27375090]

61. McCracken LM, Vowles KE, Eccleston C. Acceptance of chronic pain: component analysis and a revised assessment method. Pain 2004 Jan;107(1-2):159-166. [doi: 10.1016/j.pain.2003.10.012] [Medline: 14715402]
62. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
63. Aalen O, Veierød M, Frigessi A, Anders M, Scheel I, Skovlund E. Statistiske metoder i medisin og helsefag. Oslo, Norway: Gyldendal akademisk; 2006.
64. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
65. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to Increase Reach and Adherence of Web-Based Interventions: A Design Research Viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
66. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic Factors Influencing the Use of eHealth in People with Chronic Diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
67. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
68. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013 Aug 20;15(8):e172 [FREE Full text] [doi: 10.2196/jmir.2258] [Medline: 23963284]
69. Ahmad A, Mozelius P. Critical Factors for Human Computer Interaction of eHealth for Older Adults. In: Proceedings of the 5th International Conference on e-Society, e-Learning and e-Technologies. 2019 Presented at: 5th International Conference on e-Society, e-Learning and e-Technologies; 2019; Vienna, Austria p. 58-62. [doi: 10.1145/3312714.3312730]
70. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth Technologies: Interests, Preferences, and Concerns of Older Adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]

Abbreviations

ACT: acceptance and commitment therapy
BPI: Brief Pain Inventory
CBT: cognitive behavioral therapy
CPAQ: Chronic Pain Acceptance Questionnaire
HADS: Hospital Anxiety and Depression Scale
HRQoL: health-related quality of life
MD: mean difference
PCS: Pain Catastrophizing Scale
RCT: randomized controlled trial
SRF: Self-Regulatory Fatigue
SUS: System Usability Scale
TSD: Services for Sensitive Data

Edited by G Eysenbach; submitted 27.08.20; peer-reviewed by C Jacob, R Knoerl, H de Sola; comments to author 15.09.20; revised version received 29.09.20; accepted 02.10.20; published 23.10.20.

Please cite as:
Bostrøm K, Børøsund E, Varsi C, Eide H, Flakk Nordang E, Schreurs KMG, Waxenberg LB, Weiss KE, Morrison EJ, Cvancarova Småstuen M, Stubhaug A, Solberg Nes L
Digital Self-Management in Support of Patients Living With Chronic Pain: Feasibility Pilot Study
JMIR Form Res 2020;4(10):e23893
URL: http://formative.jmir.org/2020/10/e23893/
doi:10.2196/23893
PMID:33094734

Original Paper

Follow-Up of Cancer Patients Receiving Anti-PD-(L)1 Therapy Using an Electronic Patient-Reported Outcomes Tool (KISS): Prospective Feasibility Cohort Study

Sanna Iivanainen[1], MD, PhD; Tuomo Alanko[2], MD, PhD; Pia Vihinen[3], MD, PhD; Teemu Konkola[4], BSc; Jussi Ekstrom[4], PhD; Henri Virtanen[4], BSc; Jussi Koivunen[1], MD, PhD

[1]Department of Oncology and Radiotherapy, Oulu University Hospital, Oulu, Finland

[2]Docrates Cancer Center, Helsinki, Finland

[3]Development Unit, Hospital District of South-West Finland, Turku, Finland

[4]Kaiku Health Oy, Helsinki, Finland

Corresponding Author:
Jussi Koivunen, MD, PhD
Department of Oncology and Radiotherapy
Oulu University Hospital
PB 22
90029 OYS
Oulu
Finland
Phone: 358 83153789
Email: jussi.koivunen@ppshp.fi

Abstract

Background: Immune checkpoint inhibitors (ICIs) have become a standard of care for various tumor types. Their unique spectrum of side effects demands continuous and long-lasting assessment of symptoms. Electronic patient-reported outcome (ePRO) follow-up has been shown to improve survival and quality of life of cancer patients treated with chemotherapy.

Objective: This study aimed to investigate whether ePRO follow-up of cancer patients treated with ICIs is feasible. The study analyzed (1) the variety of patient reported symptoms, (2) etiology of alerts, (3) symptom correlations, and (4) patient compliance.

Methods: In this prospective, one-arm, multi-institutional study, we recruited adult cancer patients whose advanced cancer was treated with anti-programmed cell death protein 1 (PD)- ligand (L)1 agents in outpatient settings. The ePRO tool consisted of a weekly questionnaire evaluating the presence of typical side effects, with an algorithm assessing the severity of the symptom according to National Cancer Institute Common Terminology Criteria for Adverse Events and an urgency algorithm sending alerts to the care team. A patient experience survey was conducted monthly. The patients were followed up to 6 months or until disease progression.

Results: A total of 889 symptom questionnaires was completed by 37 patients (lung cancer, n=15; melanoma, n=9; genitourinary cancer, n=9; head and neck cancer, n=4). Patients showed good adherence to ePRO follow-up. The most common grade 1 symptoms were fatigue (28%) and itching (13%), grade 2 symptoms were loss of appetite (12%) and nausea (12%), and grade 3-4 symptoms were cough (6%) and loss of appetite (4%). The most common reasons for alerts were loss of appetite and shortness of breath. In the treatment benefit analysis, positive correlations were seen between clinical benefit and itching as well as progressive disease and chest pain.

Conclusions: According to the results, ePRO follow-up of cancer patients receiving ICIs is feasible. ePROs capture a wide range of symptoms. Some symptoms correlate to treatment benefit, suggesting that individual prediction models could be generated.

Trial Registration: Clinical Trials Register, NCT3928938; https://clinicaltrials.gov/ct2/show/NCT03928938

(JMIR Form Res 2020;4(10):e17898) doi:10.2196/17898

KEYWORDS

ePRO; immune checkpoint inhibitors; symptoms; side-effects; anti-PD-(L)1 therapy

Introduction

Cancer patients experience a variety of symptoms derived from the malignancy itself as well as side effects of the given treatment. Many symptoms are left unnoticed due to factors such as limited symptom follow-up between prescheduled health care visits, nonsystematic evaluation of symptoms, and inadequate communication [1-7]. In general, worsening of symptoms indicates cancer progression or severe side effects of the treatment and is linked to poorer cancer survival [8].

Patient-reported outcomes (PROs) consist of health-related questionnaires completed by the patients themselves, which can capture symptoms and signs and their severity. Web-based reporting of PROs has many advantages compared to paper questionnaires such as reducing time to complete and overcoming geographic location limitations. Scheduled electronic patient-reported outcomes (ePROs) enable timely and continuous collection of symptoms in a cost-effective manner [9-14]. Furthermore, use of ePROs in cancer patient monitoring has shown impressive improvements in overall survival compared to standard follow-up [15,16]. In addition, ePROs can be coupled to an urgency algorithm, which sends an alert to the care unit upon report of severe or altering symptoms by a patient. This enables rapid reaction to and treatment of important medical events.

In the past 5 years, there has been significant advancement in the development of cancer immunotherapies with the introduction of immune checkpoint inhibitor (ICI) therapies such as anti-PD-(L)1 and anti-T-lymphocyte-associated protein 4 (CTLA-4) antibodies [17]. ICI therapies have become the most important medical therapies in many malignancies such as melanoma, non-small cell lung cancer, and urogenital cancers [18-27]. ICIs differ from traditional cancer therapies due to potentially severe side effects in all organs of the body and late timing of side effect occurrence [27-29]. Therefore, there is a need for comprehensive and ongoing assessment of symptoms.

Approximately 15% of patients receiving ICI monotherapies reportedly have severe grade 3-4 side effects, and about 30% have lower grade adverse events (AEs). Even life-threatening side effects can occur, but they can, in most cases, be managed with early detection, by delaying or stopping the ICI therapy, and with the initiation of immunosuppressive medication [30-32].

To our knowledge, this is the first prospective trial investigating ePROs in the follow-up of cancer patients receiving ICIs. The study aim was to investigate the feasibility of ePRO symptom follow-up and to analyze the spectrum of patient-reported symptoms, number and aetiology of urgency algorithm alerts, correlations between different symptoms and treatment benefit, and patient compliance.

Methods

Study Design and Participants

KISS was an investigator-initiated, multicenter, prospective, one-arm study, which was undertaken in 3 multidisease cancer centers in Finland. Patients were recruited during routine doctors' appointments at study centers by study doctors. The inclusion criteria included advanced cancer to be treated with anti-PD-(L)1 in outpatient settings, initiation of anti-PD-(L)1 therapy had occurred ≤2 weeks prior to study recruitment, age ≥18 years, Eastern Cooperative Oncology Group ≤3, and availability of internet access and email. Baseline information such as basic laboratory values, age, and gender were collected from electronic health care records. After providing written informed consent, study patients received a short (5-15 minutes) instruction on how to use the Kaiku software by a study physician. At the initiation of the treatment phase (within 0-2 weeks from the first anti-PD-(L)1 infusion) and weekly thereafter until treatment discontinuation or 6 months of follow-up, patients received an email notification to complete the baseline electronic symptom questionnaire of 17 questions. If a weekly symptom questionnaire was not completed on the day of email receipt, daily email reminders were sent for 6 days. In addition, patients were asked to fill in a monthly electronic patient experience survey until treatment discontinuation or 6 months of follow-up. The use of the ePRO tool was free of charge for the patients and study centers. Online technical support by Kaiku Health for the users was available from 8 am to 4 pm Monday to Friday. The investigators evaluated the treatment response according to Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria at 8-10 weeks after treatment initiation. Clinical benefit rate was selected as a benefit measure instead of objective response rate since (1) we had a small number of study subjects, (2) responses were analyzed only up to 12 weeks from inclusion, and (3) the correlation between clinical benefit and objective response rate is not as clearly defined with immunotherapies as with traditional cancer medications.

According to the protocol, study results were analyzed when the last included patient had 12 weeks of follow-up. The major endpoints of the study included (1) patient-reported symptoms and their severity; (2) number of triggered alerts by the ePRO tool and their correlation to treatment side effects, cancer progression, other medical events, or survival; (3) correlations between different symptoms and treatment side effects, cancer progression, other medical events, or survival; and (4) patient compliance using the patient experience survey and response rates to symptom questionnaires. Sample size was based on the estimation that 15% of patients receiving ICI monotherapies will experience severe (grade 3-4) side effects and about 30% will experience lower grade AE. In a 40-patient cohort, 3-6 patients will experience a severe immune-related AE. It was estimated that the expected study population is sufficient to evaluate the feasibility of the symptom questionnaire in detecting severe AEs. Questionnaires from several timepoints were estimated to be collected from 90% of the study population (~35 patients), which would enable a more comprehensive assessment of feasibility, patient experience, and correlation of ePRO changes to treatment response and survival.

All data collection was carried out according to national legislation and under permit from the medical director of each research center. The study was approved by the Pohjois-Pohjanmaan sairaanhoitopiiri (PPSHP) ethics committee (number 9/2017), Valvira (number 361), and Oulu University

Hospital Ethics Committee (9/2017). The study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines.

ePRO Follow-Up

The Kaiku Health ePRO tool is a web-based solution scaled to be used easily on smartphones and home computers. The Kaiku Health immune-oncology module designed for the study consists of 17 questions. The symptoms selected for the Kaiku Health symptom tracking tool for cancer immunotherapy are based on the most common AEs that have occurred during clinical trials of anti-PD-1, anti-PD-L1, and anti-CTLA-4 monotherapies. The symptoms tracked by the instrument are potential signs and symptoms of immune-related AEs. The symptom selection was based on publications from the following clinical trials: CheckMate 017 (NCT01642004), CheckMate 026 (NCT02041533), CheckMate 057 (NCT01673867), CheckMate 066 (NCT01721772), CheckMate 067 (NCT01844505), KEYNOTE-010 (NCT01905657), and OAK (NCT02008227). Food and Drug Administration labels for nivolumab, pembrolizumab, and atezolizumab were also used in the symptom selection for the instrument. The questions for each symptom in the instrument were developed based on the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) register by converting the description of a grading into patient-friendly language. Any criterion that would be impossible for patients to report has been excluded from the available questions. Developing the symptom questionnaire in this manner enabled self-reporting by patients and development of an algorithm that provides an assessment and approximation of the severity of each symptom according to NCI-CTCAE criteria. NCI-CTCAE grades the symptoms from 0 to 4: no (0), mild (1), moderate (2), severe (3), and life-threatening (4).

Questions assess the presence of blood in stool, blood in urine, blurred vision, chest pain, cough, loss of appetite, diarrhea, dizziness, fatigue, fever, headache, itching, nausea, pain in joints, rash, shortness of breath, stomach pain, and vomiting. Besides recording the presence of a symptom, a severity algorithm that grades the symptom according to NCI-CTCAE was applied. The severity algorithm triggered an email alert to the study physician of the care unit based on preset limits (presence of a grade 3 or higher symptom or increase in symptom severity from grade 0 to 2). The patients were informed that the care unit would react to the alerts promptly within 3 days; thus, the ePRO follow-up was intended only for nonurgent communication, and in urgent matters, patients were advised to contact emergency care.

Patient Experience Survey

Study participants were requested to reply to a monthly patient experience survey. The patient experience survey consisted of 6 yes/no or multiple-choice questions. The survey was developed by the investigators for the study and has not been previously validated.

Statistical Analysis

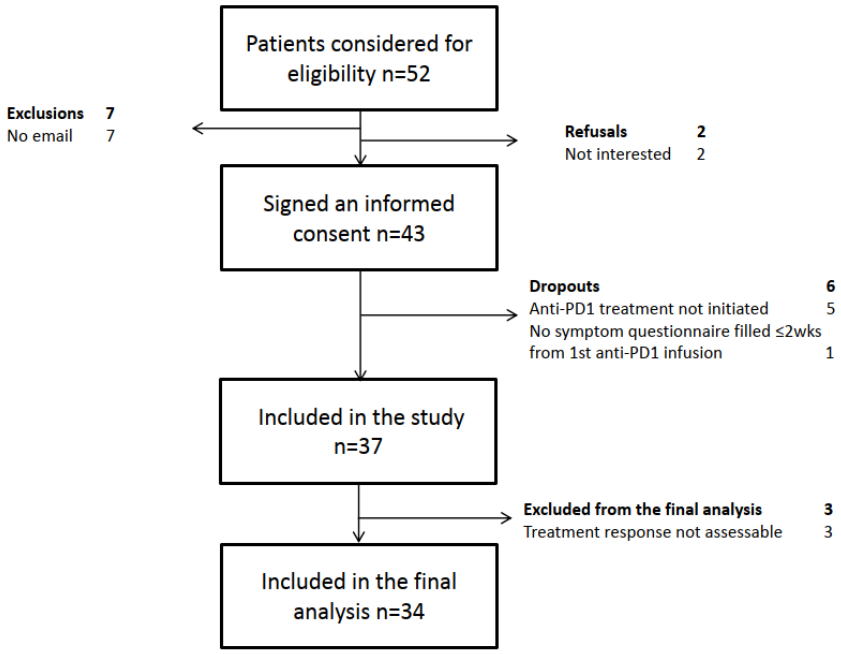
The analysis was carried out when the last patient included had 12 weeks of follow-up data available. Correlations of different patient-reported symptoms were analyzed using heat maps with Pearson product-moment correlation. In the heat map analysis, the intensity of the color signifies the level of correlation: red, negative correlation; blue, positive correlation. In other words, a large effect correlation was defined as 0.5; medium as 0.3, and small as 0.1 (absolute values).

Results

Study Accrual and Patient Characteristics

Patient recruitment took place between June 2017 and March 2019, and the last study patient visit was in June 2019. Anticipated recruitment for the study was 40 patients in 12 months, but due to a slow recruiting pace, the period was extended. Informed consent was provided by 43 patients, and analysis was limited to a total of 37 patients who had anti-PD(L)-1 therapy initiated and answered at least 2 symptom questionnaires (baseline and one following; Figure 1). No technical issues nor security breaches related to the web-based tool occurred during the study period.

Figure 1. Flowchart of patient accrual and analysis.



The median age of the study participants was 62 years (range 32-80 years). The majority of patients were male (27/37, 73%), and 5 patients had a history of an autoimmune disease, with hypothyreosis (4/5, 80%) being the most common. Tumor types included lung cancer (15/37, 41%), melanoma (9/37, 24%), genitourinary (GU) cancer (9/37, 24%), and head and neck cancer (4/37, 11%), and 28 (28/37, 76%) patients had stage IV disease (Table 1).

Table 1. Patient demographics.

Characteristics	Results
Age (years), median	61.7
Gender, n (%)	
Male	27 (73)
Female	10 (27)
Autoimmune disease, n (%)	
Yes	5 (14)
No	32 (87)
Tumor type, n (%)	
Melanoma	9 (24)
Lung cancer	15 (41)
Genitourinary cancer	9 (24)
Head and neck	4 (11)
Stage at diagnosis, n (%)	
Stage III	9 (24)
Stage IV	28 (76)
Eastern Cooperative Oncology Group (ECOG), n (%)	
0	20 (54)
1	15 (41)
2	2 (5)

Patient-Reported Symptoms and Alerts

During the study, 889 completed symptom questionnaires were registered. The range of answered questionnaires was 0.583-1.27 per patient per week, with high response rates throughout the complete follow-up period up to 24 weeks (Table 2).

Table 2. Average number of answered symptom questionnaires completed per patient per week, up to 24 weeks.

Week	Number of questionnaires per patient, mean
1	1.27
2	0.882
3	1.14
4	0.861
5	1.06
6	0.833
7	0.833
8	0.861
9	0.765
10	0.842
11	0.882
12	0.748
13	0.991
14	0.824
15	0.742
16	0.707
17	0.704
18	0.719
19	0.733
20	0.826
21	0.83
22	0.611
23	0.583
24	0.798

During the first 12 weeks of ePRO follow-up, the most common grade 1-2 symptoms were fatigue (346/889, 39%), cough (187/889, 21%), pain in joints (160/889, 18%), itching (151/889, 17%), loss of appetite (151/889, 17%), nausea (151/889, 17%), and shortness of breath (133/889, 15%). The most common grade 3-4 symptoms were cough (53/889, 6%), loss of appetite (36/889, 4%), and nausea (36/889, 4%). None of the patients (0/37) reported blood in stool or hematuria (Table 3).

Table 3. Distribution of the severity of the reported symptoms according to all the answered symptom questionnaires (n=889) in weeks 1-12.

Symptom	Grade 0, %	Grade 1, %	Grade 2, %	Grades 3-4, %
Blood in stool	100	0	0	0
Blurred vision	96	0	4	0
Chest pain	94	4	1	1
Cough	74	12	9	6
Diarrhea	96	3	1	0
Dizziness	92	6	2	0
Fatigue	60	28	11	1
Fever	95	5	0	0
Headache	87	10	2	0
Hematuria	100	0	0	0
Itching	83	13	4	1
Loss of appetite	79	5	12	4
Nausea	94	5	12	4
Pain in joints	81	12	6	2
Rash	88	9	1	1
Shortness of breath	83	8	7	2
Stomach pain	94	3	2	1
Vomiting	98	2	0	0

Of the 391 answered symptom questionnaires during the first 12 weeks, the ePRO tool triggered 67 (67/391, 17.1%) alerts. The most common reasons for alerts were loss of appetite, shortness of breath, pain in joints, blurred vision, and cough. The treating physicians were asked to evaluate the etiology of alerts by grading them to cancer, treatment, or unclear categories. Unclear reasons were the most common cause of alerts (38/67, 57%), followed by treatment (21/67, 31%) and cancer (8/67, 11%; Table 4).

Table 4. Etiology of the symptom questionnaire alerts (n=67).

Characteristics	n (%)
Etiology	
Unclear	38 (57)
Treatment	21 (31)
Cancer	8 (11)
By symptom	
Loss of appetite	32 (48)
Shortness of breath	31 (46)
Pain in joints	21 (31)
Blurred vision	17 (25)
Cough	16 (24)
Fatigue	15 (22)
Itching	12 (18)
Chest pain	9 (13)
Headache	8 (12)
Stomach pain	6 (9)
Rash	6 (9)
Nausea	5 (8)
Diarrhea	3 (5)
Dizziness	3 (5)

Patient Compliance

Patient compliance was assessed every 4 weeks based on the electronic patient experience survey provided through Kaiku software. During the first 12 weeks, 31 patients replied to the survey, and analysis was limited to these. All the patients replied that using the Kaiku software was easy or very easy, and only 1 of 6 patients reported that they needed assistance using the software. Over 90% of the patients (29/31, 94%) reported that the questions were understandable. In addition, 90% of the patients (28/31, 90%) felt that the Kaiku ePRO follow-up improved their cancer care, and 95% (29/31) said they would recommend using it in the follow-up of cancer patients (Table 5).

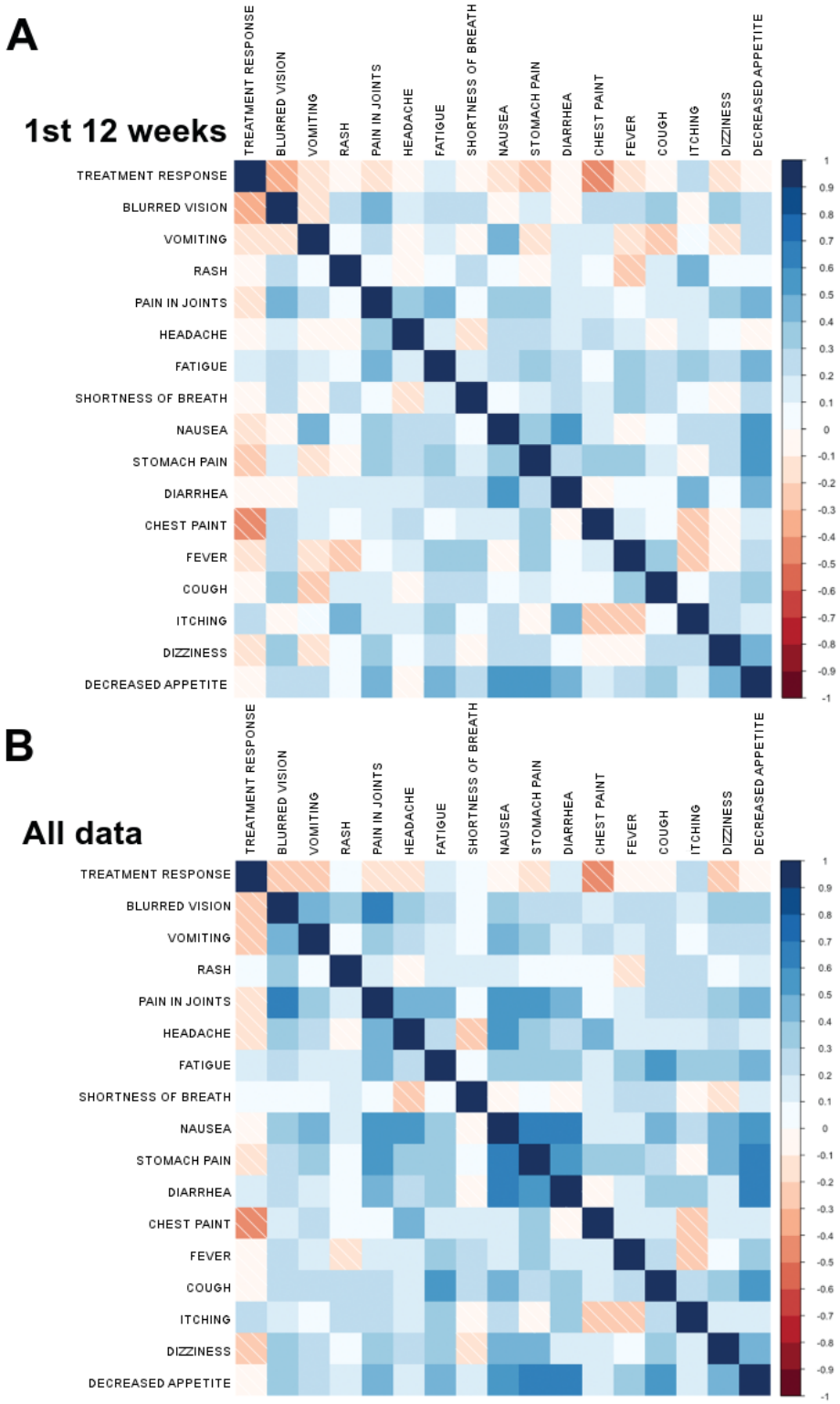
Table 5. Kaiku Health patient experience survey results during the first 12 weeks of follow-up (n=31).

Survey questions	n (%)
How easy or difficult is the use of the Kaiku Health application?	
Very easy	15 (48)
Easy	16 (52)
Difficult	0
Very difficult	0
I cannot say	0
Have you needed the help of another person to use the Kaiku Health application, not taking into account the training that you received at the health care unit?	
Yes	5 (16)
No	26 (84)
Were the questions in the symptom questionnaire in the Kaiku Health application understandable?	
Totally agree	21 (68)
Partly agree	8 (26)
Partly disagree	2 (7)
Totally disagree	0
I cannot say	0
Do you think that the use of the Kaiku Health application will improve the follow-up of your cancer treatment (compared to a situation where the application would not have been used)?	
Yes	28 (90)
No	3 (10)
I cannot say	0
Have you benefited from using the Kaiku Health application?	
Yes	19 (61)
No	1 (3)
I cannot say	11 (36)
Would you recommend the use of the Kaiku Health application in cancer care follow-up?	
Yes	29 (94)
No	0
I cannot say	2 (7)

Correlations Between Patient-Reported Symptoms and Treatment Benefit

Correlations between ePRO-collected symptoms were analyzed using heat maps. According to the results, the symptom correlations during the first 12 weeks and beyond were very similar (Figure 2). During the first 12 weeks, large positive correlations were seen between nausea, diarrhea, loss of appetite, and vomiting; stomach pain and decreased appetite; and rash and itching. Only small negative correlations were detected between cough and vomiting, itching and chest pain, and itching and fever (Figure 2A).

Figure 2. Correlation analysis between different symptoms and treatment benefit (complete response, partial response, or stable disease as a best response) using heat maps during the (A) first 12 weeks of follow-up and (B) entire study. The color intensity signifies the correlation strength (0.5, large effect; 0.3, medium effect; 0.1, small effect): red, negative correlation; blue, positive correlation.



Of the 37 patients, 34 were evaluated for objective treatment response (RECIST 1.1) by the investigators and included in the treatment benefit analysis. Of the 34 patients, 22 (65%) patients had complete response (CR), partial response (PR), or stable disease (SD) as the best response, while 12 (12/34, 35%) patients had progressive disease (PD). The heat map analysis suggested a small positive correlation between clinical benefit (CR/PR/SD) and itching (0.23 for the first 12 weeks, Figure 2A; 0.25 for all data, Figure 2B) and medium correlations between PD and chest pain (–0.41 for the first 12 weeks, Figure 2A; –0.47 for all data, Figure 2B). We further analyzed symptom progression and severity for itching and chest pain. During the first 12 weeks, 15-23% of the patients with clinical benefit reported itching, while the rate was much lower for patients with PD (0-14%;

Figure 3). Furthermore, the average grade was much higher for patients with clinical benefit (weeks 1-12, 0.26-0.37; all 12 weeks, 1.18) compared to patients with PD (weeks 1-12, 0-0.17; all 12 weeks, 0.75; Table 6). For the complete follow-up period, most of the patients with clinical benefit had itching (14/22, 64%), while this was much lower for patients with PD (4/12, 33%; Figure 3). The severity of itching for the patients with clinical benefit was mainly low grade (grade 1: 6/22, 27%; grade 2: 4/22, 18%; Figure 3). During the complete follow-up period, chest pain was much more common in patients with PD (7/12, 58%) than in the patients with clinical benefit (4/22, 18%; Figure 4). In the first 12 weeks, patients with PD had a tendency for gradually increasing average grades for chest pain; conversely, a continuing decrease in the average grade was seen for patients who responded to the therapy (Table 6).

Figure 3. Distribution of the symptom grades reported on the symptom questionnaires during the first 12 weeks for itching for (A) all patients, (B) patients with complete response (CR)/partial response (PR)/stable disease (SD), and (C) patients with progressive disease (PD).

Table 6. Average grade of itching and chest pain reported by patients.

Weeks	Itching			Chest pain		
	Entire sample (n=34)	CR[a]/PR[b]/SD[c] (n=22)	PD[d] (n=12)	Entire sample (n=34)	CR/PR/SD (n=22)	PD (n=12)
Weeks 1-2	0.16	0.26	0.00	0.06	0.09	0.04
Weeks 3-4	0.24	0.29	0.17	0.11	0.11	0.13
Weeks 5-6	0.22	0.30	0.05	0.06	0.04	0.11
Weeks 7-8	0.28	0.37	0.14	0.13	0.03	0.33
Weeks 9-10	0.20	0.30	0.00	0.07	0.00	0.22
Weeks 11-12	0.24	0.33	0.00	0.15	0.03	0.44
All 12 weeks	0.97	1.18	0.75	0.62	0.32	1.33

[a]CR: complete response.

[b]PR: partial response.

[c]SD: stable disease.

[d]PD: progressive disease.

Figure 4. Distribution of the symptom grades reported on the symptom questionnaires during the first 12 weeks for chest pain for (A) all patients, (B) patients with complete response (CR)/partial response (PR)/stable disease (SD), and (C) patients with progressive disease (PD).

Discussion

According to previous studies, ePRO follow-up has improved survival and quality of life compared to routine surveillance when used with cancer patients receiving chemotherapy and lung cancer patients treated with curative intention [15,16]. However, the ePRO approach remains virtually unstudied in the context of cancer immunotherapies [33]. For optimal follow-up of patients receiving ICIs, there is a need for comprehensive assessment, grading, and long-term surveillance of symptoms. ePROs could provide a cost-effective follow-up tool to meet these 3 requirements. We previously reported a retrospective pilot study of ePRO follow-up of cancer patients treated with ICIs [34]. To our knowledge, this study is the first prospective clinical trial investigating ePRO follow-up of cancer patients treated with anti-PD-(L)1 therapies.

In this study, we used an ePRO module with 17 questions and an algorithm grading the PROs according to NCI-CTCAE. The questionnaire was designed specifically for patients receiving ICIs based on the published side effect profile of these agents. The symptom variety based on patient reporting and the grading algorithm performed well, and the symptom data followed closely what has been reported in clinical trials investigating ICIs. A recent meta-analysis with more than 20,000 patients suggested that fatigue (18%), itching (11%), and diarrhea (9%) are the most common AEs reported in patients treated with anti-PD-(L)1 agents [35]. The incidence of AEs in clinical trials are generally lower than in our study, which might be related to better capture of patient-reported low-grade symptoms, which are often overlooked in physician-based AE reporting in clinical trials [36-42].

In the present study, the symptom questionnaire was also coupled to an urgency algorithm, which generated alerts in 17% of the answered questionnaires during the first 12 weeks. Loss of appetite and fatigue were among the most common symptoms generating alerts. These symptoms very rarely alter the cancer treatment, and symptomatic treatments are scarce. Furthermore, physicians determined that most of the alerts were caused by unclear reasons, which is probably related to the high frequency of symptoms with unclear etiology. Fine-tuning of the alerts to focus not only on the symptom grade but also the nature of the symptom could lower the number of alerts and staff workload without sacrificing the performance of ePROs.

Patient adherence to and experience with ePRO follow-up was found to be very good in this study. The patients were requested by email to complete symptom questionnaires weekly, and the number of completed questionnaires was very close to one per patient per week for the first 12 weeks. Based on the patient experience surveys, the system was easy to use, and patients felt that ePRO follow-up improved their cancer care, which is in line with previous studies [40,41].

Our previous retrospective study with patients treated with ICIs suggested that some ePRO-reported dermatological, gastrointestinal (GI), and pulmonary symptoms co-occur [34]. Similarly, we saw large positive correlations between treatment response and GI symptoms as well as between treatment response and dermatological symptoms. Furthermore, the data showed small negative correlations between pulmonary symptoms and some GI symptoms and between itching, pulmonary symptoms, and fever. In our previous retrospective study, which did not include data on the treatment responses, we generated a hypothesis that GI and skin symptoms might be related to immune activation and treatment benefit, while pulmonary symptoms could signal tumor progression. Since this study also included data on treatment benefit, it enabled us to investigate our hypothesis. The results showed that there was a small positive correlation between treatment benefit (CR/PR/SD) and itching (ePRO) and between PD and chest pain (ePRO). Previous studies have linked autoimmune skin toxicity (rash) to PD-1 agent benefit [43-45]. Our results are hypothesis-generating while suggesting that ePRO-collected symptom data can mimic physician-assessed symptoms and correlate with treatment benefit. Furthermore, compared to physician-based AE reporting, it is possible that ePROs enable enhanced capturing of low-grade AEs without visible presentation such as itching and therefore facilitate predicting clinical treatment benefit.

ePROs enable cost-effective capture of symptoms and their change over long periods [46]. Changes over time might better predict treatment side effects and benefit than just a single presentation of a symptom. Furthermore, data from this study showed that early (in the first 12 weeks) changes in symptoms correlate with treatment benefit as well as symptoms from the whole follow-up period. This further highlights the possibility that early changes in symptoms predict outcomes. Large-scale symptom data coupled with treatment benefit and side effects could be used to build prediction models using artificial intelligence methods. These models could predict an individual's risk for symptom development, treatment-related side effects, and treatment benefit.

Our study has some limitations. The sample size is small (n=37); however, the size is typical for feasibility studies. The small sample size prevents us from making strong generalizations based on the data. The one-arm design of the study precludes comparison of the effectiveness of the intervention. However, we feel that our study is important since it lays the groundwork for future studies on the topic.

In conclusion, this study is the first reported prospective clinical trial investigating the use of ePROs in the follow-up of cancer patients treated with ICIs. The results of this study suggest that follow-up of cancer patients using ePROs is feasible, enabling comprehensive capturing of symptoms over long periods with good patient adherence and satisfaction. Moreover, some early patient-reported symptoms were found to correlate with treatment benefit suggesting that individual prediction models for treatment benefit could be generated.

Acknowledgments

The study was funded by Oulu University and Finnish Cancer Society. Kaiku Health employees were involved in the data acquisition and analysis.

Authors' Contributions

SI and JK contributed to the conception and design of the study; SI, JK, TA, PV, HV, TK, and JE acquired the data; TK and JE analyzed the data; and SI and JK interpreted the data. SI and JK contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

SI and PV declare that they have no competing interests. TK, JE, and HV are employees of Kaiku Health, and HV owns stock in Kaiku Health. JK and TA are advisors for Kaiku Health.

Multimedia Appendix 1
CONSORT EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 543 KB - formative_v4i10e17898_app2.pdf]

References

1. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009 Dec 02;101(23):1624-1632 [FREE Full text] [doi: 10.1093/jnci/djp386] [Medline: 19920223]
2. Gilbert JE, Howell D, King S, Sawka C, Hughes E, Angus H, et al. Quality improvement in cancer symptom assessment and control: the Provincial Palliative Care Integration Project (PPCIP). J Pain Symptom Manage 2012 Apr;43(4):663-678 [FREE Full text] [doi: 10.1016/j.jpainsymman.2011.04.028] [Medline: 22464352]
3. Henry DH, Viswanathan HN, Elkin EP, Traina S, Wade S, Cella D. Symptoms and treatment burden associated with cancer treatment: results from a cross-sectional national survey in the U.S. Support Care Cancer 2008 Jul;16(7):791-801. [doi: 10.1007/s00520-007-0380-2] [Medline: 18204940]
4. Laugsand EA, Sprangers MA, Bjordal K, Skorpen F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. Health Qual Life Outcomes 2010 Sep 21;8(1):104 [FREE Full text] [doi: 10.1186/1477-7525-8-104] [Medline: 20858248]
5. Reilly CM, Bruner DW, Mitchell SA, Minasian LM, Basch E, Dueck AC, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer 2013 Jun 12;21(6):1525-1550 [FREE Full text] [doi: 10.1007/s00520-012-1688-0] [Medline: 23314601]
6. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008 Mar 4;17(2):179-193. [doi: 10.1007/s11136-007-9295-0] [Medline: 18175207]
7. Velikova G, Keding A, Harley C, Cocks K, Booth L, Smith AB, et al. Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: secondary outcomes of a randomised controlled trial. Eur J Cancer 2010 Sep;46(13):2381-2388. [doi: 10.1016/j.ejca.2010.04.030] [Medline: 20570138]
8. Trajkovic-Vidakovic M, de Graeff A, Voest EE, Teunissen SC. Symptoms tell it all: a systematic review of the value of symptom assessment to predict survival in advanced cancer patients. Crit Rev Oncol Hematol 2012 Oct;84(1):130-148. [doi: 10.1016/j.critrevonc.2012.02.011] [Medline: 22465016]
9. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012 Jul;62(5):337-347 [FREE Full text] [doi: 10.3322/caac.21150] [Medline: 22811342]
10. Cleeland CS, Wang XS, Shi Q, Mendoza TR, Wright SL, Berry MD, et al. Automated Symptom Alerts Reduce Postoperative Symptom Severity After Cancer Surgery: A Randomized Controlled Clinical Trial. JCO 2011 Mar 10;29(8):994-1000. [doi: 10.1200/jco.2010.29.8315]
11. Holch P, Warrington L, Bamforth L, Keding A, Ziegler L, Absolom K, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment. Ann Oncol 2017 Sep 01;28(9):2305-2311 [FREE Full text] [doi: 10.1093/annonc/mdx317] [Medline: 28911065]
12. Jensen RE, Snyder CF, Abernethy AP, Basch E, Potosky AL, Roberts AC, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. J Oncol Pract 2014 Jul;10(4):e215-e222 [FREE Full text] [doi: 10.1200/JOP.2013.001067] [Medline: 24301843]
13. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
14. Pakhomov S, Jacobsen S, Chute C, Roger V. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care 2008 Aug;14(8):530-539 [FREE Full text] [Medline: 18690769]
15. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
16. Denis F, Yossi S, Septans A, Charron A, Voog E, Dupuis O, et al. Improving Survival in Patients Treated for a Lung Cancer Using Self-Evaluated Symptoms Reported Through a Web Application. American Journal of Clinical Oncology 2017;40(5):464-469. [doi: 10.1097/coc.0000000000000189]
17. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. N Engl J Med 2012 Jun 28;366(26):2443-2454. [doi: 10.1056/nejmoa1200690]

18. Bellmunt J, de Wit R, Vaughn DJ, Fradet Y, Lee J, Fong L, et al. Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. N Engl J Med 2017 Mar 16;376(11):1015-1026. [doi: 10.1056/nejmoa1613683]
19. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. N Engl J Med 2015 Oct 22;373(17):1627-1639. [doi: 10.1056/nejmoa1507643]
20. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. N Engl J Med 2015 Jul 09;373(2):123-135. [doi: 10.1056/nejmoa1504627]
21. Herbst RS, Baas P, Kim D, Felip E, Pérez-Gracia JL, Han J, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. The Lancet 2016 Apr;387(10027):1540-1550. [doi: 10.1016/s0140-6736(15)01281-7]
22. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. N Engl J Med 2015 Nov 05;373(19):1803-1813. [doi: 10.1056/nejmoa1510665]
23. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N Engl J Med 2016 Nov 10;375(19):1823-1833. [doi: 10.1056/nejmoa1606774]
24. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. The Lancet 2017 Jan;389(10066):255-265. [doi: 10.1016/s0140-6736(16)32517-x]
25. Schachter J, Ribas A, Long GV, Arance A, Grob J, Mortier L, et al. Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006). The Lancet 2017 Oct;390(10105):1853-1862. [doi: 10.1016/s0140-6736(17)31601-x]
26. Robert C, Schachter J, Long GV, Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. N Engl J Med 2015 Jun 25;372(26):2521-2532. [doi: 10.1056/nejmoa1503093]
27. Weber JS, Hodi FS, Wolchok JD, Topalian SL, Schadendorf D, Larkin J, et al. Safety Profile of Nivolumab Monotherapy: A Pooled Analysis of Patients With Advanced Melanoma. JCO 2017 Mar 01;35(7):785-792. [doi: 10.1200/jco.2015.66.1389]
28. Li H, Ma W, Yoneda KY, Moore EH, Zhang Y, Pu LLQ, et al. Severe nivolumab-induced pneumonitis preceding durable clinical remission in a patient with refractory, metastatic lung squamous cell cancer: a case report. J Hematol Oncol 2017 Feb 28;10(1):64 [FREE Full text] [doi: 10.1186/s13045-017-0433-z] [Medline: 28245875]
29. Topalian SL, Sznol M, McDermott DF, Kluger HM, Carvajal RD, Sharfman WH, et al. Survival, Durable Tumor Remission, and Long-Term Safety in Patients With Advanced Melanoma Receiving Nivolumab. JCO 2014 Apr 01;32(10):1020-1030. [doi: 10.1200/jco.2013.53.0105]
30. Haanen J, Carbonnel F, Robert C, Kerr K, Peters S, Larkin J, ESMO Guidelines Committee. Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017 Jul 01;28(suppl_4):iv119-iv142 [FREE Full text] [doi: 10.1093/annonc/mdx225] [Medline: 28881921]
31. Puzanov I, Diab A, Abdallah K, Bingham CO, Brogdon C, Dadu R, Society for Immunotherapy of Cancer Toxicity Management Working Group. Managing toxicities associated with immune checkpoint inhibitors: consensus recommendations from the Society for Immunotherapy of Cancer (SITC) Toxicity Management Working Group. J Immunother Cancer 2017 Nov 21;5(1):95 [FREE Full text] [doi: 10.1186/s40425-017-0300-z] [Medline: 29162153]
32. Spain L, Diem S, Larkin J. Management of toxicities of immune checkpoint inhibitors. Cancer Treat Rev 2016 Mar;44:51-60. [doi: 10.1016/j.ctrv.2016.02.001] [Medline: 26874776]
33. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized Trial Comparing a Web-Mediated Follow-up With Routine Surveillance in Lung Cancer Patients. J Natl Cancer Inst 2017 Sep 01;109(9):029. [doi: 10.1093/jnci/djx029] [Medline: 28423407]
34. Iivanainen S, Alanko T, Peltola K, Konkola T, Ekström J, Virtanen H, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar 21;145(3):765-774 [FREE Full text] [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
35. Wang DY, Salem J, Cohen JV, Chandra S, Menzer C, Ye F, et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. JAMA Oncol 2018 Dec 01;4(12):1721-1728 [FREE Full text] [doi: 10.1001/jamaoncol.2018.3923] [Medline: 30242316]
36. Basch E, Dueck AC, Rogak LJ, Minasian LM, Kelly WK, O'Mara AM, et al. Feasibility Assessment of Patient Reporting of Symptomatic Adverse Events in Multicenter Cancer Clinical Trials. JAMA Oncol 2017 Aug 01;3(8):1043-1050 [FREE Full text] [doi: 10.1001/jamaoncol.2016.6749] [Medline: 28208174]
37. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of Patient-Reported Outcomes in Routine Medical Care. Am Soc Clin Oncol Educ Book 2018 May 23;38:122-134. [doi: 10.1200/EDBK_200383] [Medline: 30231381]
38. Basch E, Rogak LJ, Dueck AC. Methods for Implementing and Reporting Patient-reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. Clin Ther 2016 Apr;38(4):821-830 [FREE Full text] [doi: 10.1016/j.clinthera.2016.03.011] [Medline: 27045992]
39. Basch E, Dueck AC, Rogak LJ, Mitchell SA, Minasian LM, Denicoff AM, et al. Feasibility of Implementing the Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events in a Multicenter Trial: NCCTG N1048. JCO 2018 Nov 01;36(31):3120-3125. [doi: 10.1200/jco.2018.78.8620]

40. Wright E, Selby P, Crawford M, Gillibrand A, Johnston C, Perren T, et al. Feasibility and Compliance of Automated Measurement of Quality of Life in Oncology Practice. JCO 2003 Jan 15;21(2):374-382. [doi: 10.1200/jco.2003.11.044]
41. Velikova G, Wright EP, Smith AB, Cull A, Gould A, Forman D, et al. Automated Collection of Quality-of-Life Data: A Comparison of Paper and Computer Touch-Screen Questionnaires. JCO 1999 Mar;17(3):998-998. [doi: 10.1200/jco.1999.17.3.998]
42. Basch E, Artz D, Dulko D, Scher K, Sabbatini P, Hensley M, et al. Patient Online Self-Reporting of Toxicity Symptoms During Chemotherapy. JCO 2005 May 20;23(15):3552-3561. [doi: 10.1200/jco.2005.04.275]
43. Freeman-Keller M, Kim Y, Cronin H, Richards A, Gibney G, Weber JS. Nivolumab in Resected and Unresectable Metastatic Melanoma: Characteristics of Immune-Related Adverse Events and Association with Outcomes. Clinical Cancer Research 2015 Oct 07;22(4):886-894. [doi: 10.1158/1078-0432.ccr-15-1136]
44. Sanlorenzo M, Vujic I, Daud A, Algazi A, Gubens M, Luna SA, et al. Pembrolizumab Cutaneous Adverse Events and Their Association With Disease Progression. JAMA Dermatol 2015 Nov 01;151(11):1206-1212 [FREE Full text] [doi: 10.1001/jamadermatol.2015.1916] [Medline: 26222619]
45. Berner F, Bomze D, Diem S, Ali OH, Fässler M, Ring S, et al. Association of Checkpoint Inhibitor-Induced Toxic Effects With Shared Cancer and Tissue Antigens in Non-Small Cell Lung Cancer. JAMA Oncol 2019 Jul 01;5(7):1043-1047 [FREE Full text] [doi: 10.1001/jamaoncol.2019.0402] [Medline: 31021392]
46. Lizée T, Basch E, Trémolières P, Voog E, Domont J, Peyraga G, et al. Cost-Effectiveness of Web-Based Patient-Reported Outcome Surveillance in Patients With Lung Cancer. J Thorac Oncol 2019 Jun;14(6):1012-1020 [FREE Full text] [doi: 10.1016/j.jtho.2019.02.005] [Medline: 30776447]

Abbreviations

AE: adverse event
CTLA-4: cytotoxic T-lymphocyte-associated protein 4
CR: complete response
ECOG: Eastern Cooperative Oncology Group
ePRO: electronic patient-reported outcome
GI: gastrointestinal
ICI: immune checkpoint inhibitor
NCI-CTCAE: National Cancer Institute Common Terminology Criteria for Adverse Events
PD: progressive disease
PD-(L)1: protein 1-ligand 1
PR: partial response
PRO: patient-reported outcome
RECIST: Response Evaluation Criteria in Solid Tumors
SD: stable disease

Edited by G Eysenbach; submitted 20.01.20; peer-reviewed by D Howell, P Innominato, R Cruz Martínez, J Wilkinson; comments to author 20.04.20; revised version received 12.05.20; accepted 22.09.20; published 28.10.20.

Please cite as:
Iivanainen S, Alanko T, Vihinen P, Konkola T, Ekstrom J, Virtanen H, Koivunen J
Follow-Up of Cancer Patients Receiving Anti-PD-(L)1 Therapy Using an Electronic Patient-Reported Outcomes Tool (KISS): Prospective Feasibility Cohort Study
JMIR Form Res 2020;4(10):e17898
URL: http://formative.jmir.org/2020/10/e17898/
doi:10.2196/17898
PMID:33112242

Original Paper

Adaptation of a Digital Health Innovation to Prevent Relapse and Support Recovery in Youth Receiving Services for First-Episode Psychosis: Results From the Horyzons-Canada Phase 1 Study

Shalini Lal[1,2,3], BScOT, MSc, PhD; John Gleeson[4], PhD; Lysanne Rivard[2], MA, PhD; Simon D'Alfonso[5], BASc, PhD; Ridha Joober[3,6], PhD, MD; Ashok Malla[3,6], MD, FRCPC; Mario Alvarez-Jimenez[7,8], DClinPsy, MResMeth, PhD

[1]School of Rehabilitation, Faculty of Medicine, University of Montréal, Montréal, QC, Canada

[2]Youth Mental Health and Technology Lab, Health Innovation and Evaluation Hub, University of Montréal Hospital Research Centre, Montréal, QC, Canada

[3]Prevention and Early Intervention Program for Psychosis (PEPP-Montreal) and ACCESS Open Minds, Douglas Mental Health University Institute, Montréal, QC, Canada

[4]Healthy Brain and Mind Research Centre and School of Behavioural and Health Sciences, Australian Catholic University, Fitzroy, Australia

[5]School of Computing and Information Systems, The University of Melbourne, Parkville, Australia

[6]Department of Psychiatry, McGill University, Montréal, QC, Canada

[7]Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

[8]Orygen, Parkville, Australia

Corresponding Author:
Shalini Lal, BScOT, MSc, PhD
School of Rehabilitation
Faculty of Medicine
University of Montréal
C.P. 6128, succursale Centre-ville
Montréal, QC, H3C 3J7
Canada
Phone: 1 514 890 8000 ext 31581
Email: shalini.lal@umontreal.ca

Abstract

Background: Developing a digital health innovation can require a substantial amount of financial and human resource investment before it can be scaled for implementation across geographical, cultural, and health care contexts. As such, there is an increased interest in leveraging eHealth innovations developed and tested in one country or jurisdiction and using these innovations in local settings. However, limited knowledge exists on the processes needed to appropriately adapt digital health innovations to optimize their transferability across geographical, cultural, and contextual settings.

Objective: We report on the results of an adaptation study of Horyzons, a digital health innovation originally developed and tested in Australia. Horyzons is designed to prevent relapses and support recovery in young people receiving services for first-episode psychosis (FEP). The aim of this study is to assess the initial acceptability of Horyzons and adapt it in preparation for pilot testing in Canada.

Methods: This research took place in 2 specialized early intervention clinics for FEP, located in 1 urban and 1 urban-rural setting, in 2 Canadian provinces. A total of 26 participants were recruited: 15 clinicians (age range 26-56 years) and 11 patients (age range 19-37 years). Following the digital health adaptation framework developed by our team, we used a mixed methods approach, combining descriptive quantitative and qualitative methods across 3 stages of data collection (focus groups, interviews, and consultations), analysis, and adaptations.

Results: Overall, patients and clinicians appreciated the strengths-based approach and social media features of Horyzons. However, participants expressed concerns related to implementation, especially in relation to capacity (eg, site moderation, crisis management, internet speed in rural locations). They also provided suggestions for adapting content and features, for example, in relation to community resources, volume of text, universal accessibility (eg, for individuals with limitations in vision), and optimization of platform accessibility through mobile devices. Additional aspects of the innovation were flagged for adaptation during the final stages of preparing it for live implementation. These included terms of use, time zone configuration to reflect

local time and date, safety and moderation protocols, the *need help now* feature, and the list of trigger words to flag posts indicative of potential risk.

Conclusions: In the context of the COVID-19 pandemic and public health guidelines for social distancing, there is an increasing interest and need to leverage the internet and mobile technologies for delivering youth mental health services. As countries look to one another for guidance on how to navigate changing social dynamics, knowledge on how to utilize and adapt existing innovations across contexts is now more important than ever. Using a systematic approach, this study illustrates the methods, processes, results, and lessons learned on adapting a digital health innovation to enhance its local acceptability.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.8810

(JMIR Form Res 2020;4(10):e19887) doi:10.2196/19887

KEYWORDS

psychotic disorders; mental health; telemedicine; young adult; mental health services; cultural adaptation; mobile phone; e-mental health; virtual care; schizophrenia; e-health

Introduction

Digital Health Adaptations

Health service providers, community organizations, and consumers are increasingly looking toward leveraging digital health/eHealth innovations developed and tested in one part of the world (eg, country, state, region) for importation and use in their local communities. This approach can help reduce duplication of efforts and make better use of investments from the public, private, and philanthropic sectors of the community [1]. At the same time, little is known about the optimal processes involved in adapting digital health innovations when implementing them across geographical, cultural, and other types of contexts [1,2].

A *digital health (eHealth) adaptation* refers to a "systematic, purposeful, and collaborative process of making changes" to a digital health innovation to increase its "relevance and acceptability" for a local community [1]. Building on the adaptation models from the education, psychology, and technology fields, eHealth adaptation involves the consideration of factors such as language, culture, and context [1,3,4] when tailoring innovations for use in local settings. The ultimate purpose of adaptation is to increase the likelihood for an innovation to have value and impact for the population it is being adapted for. This is in alignment with the current guidelines for developing and evaluating complex interventions that highlight the importance of adapting interventions to local settings to improve their impact potential [5].

Indeed, within the broader mental health literature, previous studies have indicated that interventions originally designed for a population and adapted to the cultural and contextual needs of another population can yield better outcomes than providing those interventions to the same population without adaptation [2,6-8]. However, there are inconsistencies in terms of the extent of adaptations reported across studies [2]. There are also challenges in reviewing the effectiveness of adapted eHealth innovations given the limited consensus on guidelines, methods, and processes for adaptation and limited documentation on the adaptations undertaken by researchers when transporting or importing digital health innovations across contexts [1].

The Horyzons Platform

In a previous paper, we reported on a digital health adaptation framework and research protocol for adapting Horyzons, an innovative digital mental health intervention that was originally developed and tested in Australia, in preparation for a pilot implementation study in Canada [1]. Horyzons is a web-based application powered by the Moderated Web-Based Social Therapy (MOST) system. MOST consists of interactive and strengths-focused psychosocial interventions, web-based social networking, and clinical and peer moderation. By tailoring the therapy content to target the treatment of specific conditions and adding any required code customizations, the flexible MOST platform enables the setting up of individual sites for a variety of mental health cohorts. Horyzons is a MOST-based intervention originally designed for cohorts of young people recovering from first-episode psychosis (FEP). The original version of Horyzons was developed iteratively over 30 months following participatory design principles and user-centered methodologies by an interdisciplinary team of experts in collaboration with young people receiving specialized early intervention (SEI) services for FEP [9].

Horyzons has been tested on a sample of 20 young Australian adults for its feasibility, acceptability, utility, and safety [9,10], and more recently, it has been pilot-tested in other international contexts [11]. There is also a randomized controlled trial of Horyzons [12] that was completed in 2018 and is currently undergoing data analysis. In addition, the MOST platform has been improved and adapted to address the needs of young people and their caregivers across the diagnostic spectrum [13-17]. Further details on the Horyzons platform and its core features are provided in our research protocol detailing phase 1 of our international research program [1].

Objectives

The objectives of phase 1 were to assess the initial acceptability of Horyzons and adapt it in preparation for phase 2 pilot testing in Canada. In this paper, we report on the results of our phase 1 adaptation study.

Methods

Study Design and Setting

This study applied a mixed methods approach, combining descriptive quantitative and qualitative methods. The research took place in 2 SEI clinics for FEP, located in 1 urban and 1 small urban-rural setting in 2 Canadian provinces (Quebec and Ontario). Both programs provide a comprehensive range of services for young people diagnosed with FEP and follow best practice guidelines [18,19]. Ethics approval was first obtained from the ethics review board of the primary recruitment site before seeking and obtaining approval from the ethics review board of the secondary site.

Participants

All participants (clinicians and patients) provided written informed consent before participating in the study. Eligibility criteria for patients were as follows: diagnosed with a psychotic disorder, within their first 3 years of treatment, currently in treatment, considered to be symptomatically stable and capable of participating in focus groups as judged by their primary treating clinician, 18 years of age or older, and the ability to speak and read in English. Eligibility criteria for clinicians were psychiatrists, case managers, or other health care professionals with a minimum of 2 years of experience working in the field of SEI for FEP and regularly involved in delivering services to youths with FEP.

Adaptation Framework and Data Collection

The adaptation framework informed the data collection methods and processes. The framework is organized by 3 objectives and stages: (1) assess the initial perspectives of service users and providers of the eHealth intervention (before any modifications) following a brief orientation to the website, (2) assess the perspectives of the eHealth intervention following an extended exploration of the website, and (3) adapt the eHealth intervention on the basis of feedback from key stakeholder groups (while respecting its core therapeutic elements and principles and considering feasibility). Further details on the methods and processes of the eHealth adaptation framework are described in our previous protocol publication [1].

Data collection included a sociodemographic and technology use questionnaire, focus groups, and written feedback forms (Table 1). Sociodemographic and technology use questionnaires were used to ascertain the sociodemographic features of the participant sample (for both clinicians and patients) and to better understand their baseline access to and experience with technology.

Table 1. Summary of data collection activities.

Stage	Participants	Data collection activities
1	11 patients: 6 urban, 5 urban-rural	Sociodemographic and technology use questionnaire
	15 clinicians: 10 urban, 5 urban-rural	Exploration of beta version of the web portal and focus group discussion (90-120 min). Includes the following steps: • Assigned to individual computer stations and provided user access • Orientation to the web portal • Individual exploration of the web portal • Individual feedback forms, for example, general impressions; usefulness, safety, and support; design, layout, navigation; accessibility; and organizational capacities (the latter pertaining to clinicians only) • Group discussion
2	6 patients: 4 urban, 2 urban-rural; 4 clinicians: 3 urban, 1 urban-rural	Personal extended exploration of beta version of the site, including the following steps: Up to 120 min over a 2- to 4-week duration from a personal computer, guided by a set of instructions on activities to complete; completion of the feedback form on personal impressions and suggestions for modifications and adaptations
3	4 clinicians: 2 urban, 2 urban-rural	With clinicians: individual consultation interview and follow-up meetings to review and complete recommended adaptations
	2 patients: 1 urban, 1 urban-rural	With patients: reviewing modifications and completion of the feedback form

In alignment with our eHealth adaptation framework, data collection was conducted in 3 iterative stages by a research assistant (LR). Clinicians and patients participated in all the 3 stages in separate groups. In the first stage, all participants were invited to complete the sociodemographic and technology use questionnaire, followed by an exploration of the web portal within the context of a focus group discussion (see Multimedia Appendix 1 for the visual setups of the group discussion). During the focus group, participants were asked questions on likes and dislikes of the platform, helpfulness of the content and features, how engaging the platform is to use, factors that could reduce motivation, platform safety, thoughts on the supports offered, design and layout, ease of use, difficult-to-understand words or expressions, and suggestions for adaptations (eg, language, metaphor, ease of use, content).

The second stage consisted of personal exploration of the platform. All participants were given log-in access to the platform. They were invited to explore its content and features over a 2- to 4-week duration on their personal computer, complete a set of activities on the platform, and fill out a personal feedback form regarding their comments and suggestions for adaptations. Feedback topics included items such as appropriateness of language; likes, dislikes, facilitators, and barriers; usefulness; safety; design and ease of use; therapeutic alignment; and organizational factors. Clinicians

and patients were encouraged to explore 4 features, activities, or modules to evaluate (2 preselected by researchers, for clinicians these were preselected on the basis of their area of expertise, and 2 selected by the participants). Clinicians provided written feedback for each activity that included suggested adaptations, reasons why the adaptations were important, and general comments. Patients provided written feedback related to each activity regarding content; difficult-to-understand sentences or words; and likes, dislikes, and general comments. Patients were also asked additional questions regarding other activities they tried on the website, their comments on those other activities (eg, likes, dislikes, recommendations for change), and any additional comments or suggestions.

In the third stage, we conducted individual consultation interviews with clinicians to complete the recommended adaptations and interviews with patients to review the adaptations. This process was iterative in the sense that it involved back and forth communication with participants and revisions to arrive at a satisfactory version of the adaptations.

Data Management and Analysis

Quantitative data from the sociodemographic and technology use questionnaires were analyzed using simple descriptive statistics (counts and percentages). In terms of data management, the qualitative data from the focus group discussions were recorded and transcribed verbatim and written feedback responses from the focus group and the extended exploration phase were organized into tables. The qualitative data were managed using Atlas.ti (version 7.5.6; ATLAS.ti Scientific Software Development GmbH), which supported the coding and analysis process. First, a preliminary coding framework was developed on the basis of the research objectives and initial review and discussion of the data by SL and LR. Once the coding framework was finalized (Multimedia Appendix 2), it was systematically applied to all the data by LR. A synthesis table of the data for each of the codes was then prepared per site (including the participant ID, data source, and verbatim citation) and themes were identified on the basis of patterns within the data (ie, topic mentioned by a minimum of 3 participants). The coding was also audited by a second research assistant who was not involved in the data collection and who was asked to assign codes to the extracted data. This audit process revealed minor discrepancies, which were discussed among the research teams to arrive at a consensus. A final review of the coding was then conducted by SL.

Results

Sociodemographic Characteristics

In total, 11 patients and 15 clinicians participated in this adaptation study. More than half of the patients were males (6/11, 55%) and between the ages of 19 and 25 years (6/11, 55%). The majority of patients reported completing a college-level education or higher (8/11, 73%); engaging in school, work, or caregiving activities (8/11, 73%); an annual income of less than Can $14,999 (approximately US $10,499) (9/11, 82%); and receiving specialized services for FEP for more than 2 years (7/11, 64%; see Table 2 for additional details on the sociodemographic characteristics of the patients).

Table 2. Sociodemographic breakdown of patients (total=11; urban=6; urban-rural=5).

Characteristics	Values, n (%)
Gender	
Male	6 (55)
Female	4 (36)
Other	1 (9)
Age (years)	
19-25	6 (55)
26-37	5 (45)
Race	
White	7 (64)
Two or more races	3 (27)
Hispanic	1 (9)
Education (highest level completed)	
Less than high school	1 (9)
High school	2 (18)
College/vocational degree or diploma	5 (45)
Bachelor's degree	2 (18)
Graduate diploma	1 (9)
Length of receiving services at urban or urban-rural site	
0-6 months	4 (36)
>2 years	7 (64)
Current occupation	
Not in education, employment, or caregiving activities	3 (27)
Working	5 (45)
In school	1 (9)
In school and working	2 (18)
Annual household income, Can $ (US $)	
Under 14,999 (10,499) per year	9 (82)
15,000-29,000 (10,500-20,300) per year	1 (9)
50,000 (35,000) and above per year	1 (9)
Current living situation	
Live alone	5 (45)
Live with parent(s)/sibling(s)	5 (45)
Live with partner and children	1 (9)
Current relationship status	
Single	8 (73)
In a relationship	2 (18)
Legally married	1 (9)
Access to a smartphone	
Yes	9 (82)
No	1 (9)
Other/sometimes	1 (9)

Characteristics	Values, n (%)
Access to a home computer	
Yes	8 (73)
No	3 (27)
Use of the internet to search for mental health information, services, and support	
Once per week or less	8 (73)
Once per week	1 (9)
2 to 3 times per week	2 (18)
Use of social media to communicate with others	
Yes	9 (82)
No	2 (18)
Frequency of use of social media	
Once per week or less	2 (18)
2 to 3 times per week	2 (18)
Daily	5 (45)
N/A[a]	2 (18)
Use of SMS text messaging to communicate with others	
Yes	8 (73)
No	3 (27)
Frequency of use of SMS text messaging	
2 to 3 times per week	1 (9)
Daily	7 (64)
N/A[a]	3 (27)
Use of email to communicate with others	
Yes	9 (82)
No	2 (18)
Frequency of use of email	
Once per week or less	1 (9)
Once per week	1 (9)
2 to 3 times per week	2 (18)
Daily	5 (45)
N/A[a]	2 (18)
To what extent do you feel competent in using a computer	
Not competent	0 (0)
Somewhat not competent	0 (0)
Neutral	4 (36)
Somewhat competent	4 (36)
Very competent	3 (27)
To what extent do you feel competent in using social media to communicate with others	
Not competent	1 (9)
Somewhat not competent	0 (0)
Neutral	2 (18)
Somewhat competent	5 (45)

Characteristics	Values, n (%)
Very competent	3 (27)
To what extent do you feel competent in searching the internet for mental health information, services, and supports	
Not competent	0 (0)
Somewhat not competent	2 (18)
Neutral	4 (36)
Somewhat competent	4 (36)
Very competent	1 (9)

[a]N/A: not applicable.

In terms of patient access, use, and perceived competence with technology (Table 2), the majority (9/11, 82%) had access to a smartphone and to a home computer (8/11, 73%). Only a few patients (3/11, 27%) reported using the internet to search for mental health information, services, and supports once per week or at least two to three times per week; more than half of the patients reported using SMS text messaging (7/11, 64%) daily and social media (7/11, 64%) daily or at least two to three times per week to communicate with others. Three patients reported not using SMS text messaging and 2 patients reported not using social media for communication purposes at all. In terms of perceived competence with technology, the majority of patients reported feeling at least somewhat competent with respect to using a computer (7/11, 64%) and using social media to communicate with others (8/11, 73%). Slightly less than half of the patients reported feeling at least somewhat competent searching the internet for mental health–related content (5/11, 45%), and the remaining patients reported feeling neutral (4/11, 36%) or somewhat not competent (2/11, 18%).

With regard to clinicians, the majority were women (12/15, 80%), between the ages of 26 and 56 years, had a master's degree (9/15, 60%), and had a professional background in social work (7/15, 47%) or nursing (3/15, 20%). Their duration of time working in youth mental health services ranged from 2 to 16 years (see Table 3 for additional details on the sociodemographic characteristics of the clinicians).

In terms of clinician access, use, and perceived competence with technology (Table 3), the majority (14/15, 93%) had access to a smartphone and all had access to a home computer. A little over half of the clinicians (8/15, 53%) reported using the internet to search for work-related mental health information, services, and supports daily or at least two to three times per week. The majority reported using SMS text messaging (13/15, 87%) daily, and more than half reported using social media (10/15, 67%) daily to communicate with others. In terms of perceived competence with technology, approximately half of the clinicians (7/15, 47%) reported feeling somewhat competent when using a computer and social media, with 40% (6/15) feeling very competent using a computer and 27% (4/15) feeling very competent using social media to communicate with others. Approximately 60% (9/15) reported feeling very competent when searching the internet for mental health–related content, and the remaining 40% (6/15) indicated feeling somewhat competent with this activity.

Table 3. Sociodemographic breakdown of clinicians (total=15; urban=10; urban-rural=5).

Characteristics	Values, n (%)
Sex	
Male	3 (20)
Female	12 (80)
Age (years)	
26-40	7 (47)
42-56	8 (53)
Race	
White	14 (93)
Asian	1 (7)
Education (highest level completed)	
College/vocational degree/diploma	3 (20)
Bachelor's degree	2 (13)
Master's degree	9 (60)
Medical degree	1 (7)
Professional discipline	
Psychology	1 (7)
Psychiatry	1 (7)
Human relations	1 (7)
Occupational therapy	2 (13)
Nursing	3 (20)
Social work	7 (47)
Length of time working in youth mental health services (years)	
2-5	4 (27)
6-10	6 (40)
11-16	5 (33)
Access to a smartphone	
Yes	14 (93)
Other/sometimes	1 (7)
Access to a home computer	
Yes	15 (100)
Use of the internet to search for work-related mental health information, services, and supports	
Once per week or less	7 (47)
2 to 3 times per week	2 (13)
Daily	6 (40)
Use of social media to communicate with others	
Yes	12 (80)
No	3 (20)
Frequency of use of social media	
2 to 3 times per week	2 (13)
Daily	10 (67)
N/A[a]	3 (20)

Characteristics	Values, n (%)
Use of SMS text messaging to communicate with others	
Yes	15 (100)
Frequency of use of SMS text messaging	
Once per week or less	1 (7)
2 to 3 times per week	1 (7)
Daily	13 (87)
Use of email to communicate with others	
Yes	14 (93)
No	1 (7)
Frequency of use of email	
Once per week or less	4 (27)
2 to 3 times per week	3 (20)
Daily	7 (47)
N/A[a]	1 (7)
To what extent do you feel competent in using a computer	
Not competent	0 (0)
Somewhat not competent	1 (7)
Neutral	1 (7)
Somewhat competent	7 (47)
Very competent	6 (40)
To what extent do you feel competent in using social media to communicate with others	
Not competent	1 (7)
Somewhat not competent	2 (13)
Neutral	1 (7)
Somewhat competent	7 (47)
Very competent	4 (27)
To what extent do you feel competent in searching the internet for mental health information, services, and supports	
Not competent	0 (0)
Somewhat not competent	0 (0)
Neutral	0 (0)
Somewhat competent	6 (40)
Very competent	9 (60)

[a]N/A: not applicable.

Perspectives on the Digital Health Innovation

The results from the focus groups and written feedback from the extended explorations are organized according to the following themes: (1) appreciating the therapeutic approach and relatability of Horyzons; (2) diverging opinions on design, layout, and ease of navigation; (3) being concerned about implementation; and (4) providing suggestions for changing content and features. The following sections provide further details on each of these themes with illustrations through participant quotes. Quotes are labeled with ID codes that represent the stakeholder group (ie, patient, P, or clinician, C). Given the small sample size, we removed gender and site-level identification to protect the anonymity of the participants.

Appreciating the Therapeutic Approach and Relatability of Horyzons

Overall, the majority of patients (10/11, 90%) and all clinicians (15/15, 100%) expressed appreciation of the platform in terms of its therapeutic approach (eg, related to content, activities, social media features) and relatability for young people, including those that have experienced FEP. For example, in terms of relatability, patient participants stated:

> *The website does a nice job of outlining scenarios that are very relatable as well as steps for how to deal with these different instances.* [P1]
>
> *I enjoyed reading people's stories because it is from a trusted site and it was interesting to see stuff I could relate to from their story.* [P2]

Similarly, clinicians mentioned that the stories of characters presented on the website "provided a relatable way to deliver psychoeducation about psychosis," (C1) and that the content is relatable for youth:

> *It's more of a general site for everybody, which I think is good. And I get the impression that these are general topics that young adults and young adolescents anyway would talk about.* [C2]

Positive comments about Horyzons were made by both clinicians and patients in reference to its content, strengths-focused activities, and social media features. In terms of content, clinicians commented that the platform was informative, for example:

> *As a platform for information, I think it's fantastic and I think our client, from entry to exit, could easily benefit from this.* [C3]

Patient participants highlighted their interest in topics addressed in the website:

> *I liked the content itself. I find it interesting topics and stuff.* [P3]

Across both stakeholder groups, participants expressed appreciation for how content was presented and organized, in relation to tips, things to do on the website, and multimedia:

> *I appreciate the videos and audio tracks, not everyone wants to read a lot of material. It is easier to follow audio tracks also for certain types of exercises.* [C4]
>
> *I found the tips helpful.* [P4]
>
> *I found that since there are lots of [activities], like it can help many people.* [P5]

Clinician and patient participants also linked the therapeutic value of the platform to evidence-based psychosocial therapies such as cognitive behavioral therapy and mindfulness therapy:

> *It's like CBT! … tips to change your behav[iour], to improve your wellbeing.* [P7]
>
> *Good 'training' on affective regulation, mindfulness in an understandable and user-friendly way.* [C1]

In terms of the strengths-based approach, patients and clinicians highlighted the personal strengths card sorting activity conducted at the beginning of the website intervention:

> *I can see how this would make someone feel pretty positive about themselves … because there's so many to choose from, like 'Wow! I have way more than 5 top strengths!* [C5]

Several of the participants (patients and clinicians) reacted favorably toward the social media and community features of the platform and expressed appreciation of how they resembled mainstream social media, such as Facebook (P8; C6) and how the website includes clinical moderation with peer networking:

> *Mixing the community part with the like helping part. I think that's a good idea.* [P5]

Patient participants highlighted the value of these features in relation to accessing peer support:

> *Lets people know that other people experience similar things.* [P10]
>
> *Being able to talk to other people that have or are going through similar things.* [P11]
>
> *We can exchange with others on certain subjects/share stories.* [P7]

In addition, clinicians commented that Horyzons could be a useful resource when patients start receiving specialized services for FEP in addition to transitions between services. For example, one clinician explained how the site could help patients stay engaged in their treatment:

> *The stories and explanations are great and could…encourage newly admitted FEP clients to maintain their treatment, alliance building, etc.* [C3]

Diverging Opinions on Design, Layout, and Ease of Navigation

Participants across both groups reported contradicting opinions on the design, layout, and navigation. For instance, some appreciated the calm and neutral look of the site. One patient explained:

> *I like that there's nothing that's too distracting on it* [P3]

Others would have preferred a more colorful design or sounds to certain features for a more dynamic and interactive experience. Only 1 patient raised general issues with respect to navigating the site, in contrast to several clinicians who expressed challenges navigating information on the site. For example, clinicians expressed that there was a lot of text to navigate on the platform, that the exploratory navigation style of the platform may lead to some users disengaging from the site, and that the organization of content could be better organized:

> *It would be more useful for the content to be organized under clear headings rather than have to sift through what has been posted on the main wall.* [C7]

Being Concerned About Implementation

Clinicians raised several concerns about its implementation. These concerns pertained to patient safety, clinical moderation, and internet connectivity. For example, clinicians raised concerns about the respectful use of the site:

> *I wonder what the "café" comments and other comments from users will look like. Will they be respectful? There could be some occasional problems.* [C1]

Clinicians also voiced concerns about users witnessing others on the site going through a crisis, crisis response management,

integration of moderation into clinical workflow, and new roles and expectations (eg, in relation to web-based moderation and web-based crisis management response). Regarding the latter, some clinicians were concerned that if web-based moderation of the site became a part of their workday, it would take away from their duties as a case manager:

> *I think there is a certain amount of time that a clinician will have to look through it. So, whatever time they are spending doing that, it is time they are not actually doing case managing…personally, for me, it is not the type of work I would do with my clients.* [C7]

Clinicians also expressed concern regarding technology infrastructure. Participants from the urban-rural setting experienced navigation and uploading difficulties due to a very slow internet connection. They highlighted that having internet connectivity issues was common in their setting and something that will need to be considered in a live implementation scenario:

> *Our internet is not going to be like it is in the city… but that's something that they're gonna have to think about, right? Because it is really really different access here.* [C9]

Patients also raised concerns about access and safety. They reported that some individuals may not want to connect with others on the web, "they might prefer to deal with their problems more privately" (P1); might have limited access to the technology needed to participate on the site, "access to the internet/computer capable of using the site" (P11); or might be deterred from using the site due to their psychosis symptoms, "fear of being tracked about what is being said or done (paranoia for some youth)" (P11) or due to concerns "about online bullying" (P11). They offered several strategies to promote engagement with the platform, including having its use:

> *Encouraged by their doctor or peers.* [P1]
>
> *Making the site cell phone accessible.* [P11]
>
> *Putting stories on there about successful recovery (normal people and celebrities alike).* [P11]

Recommending Changes to Content and Features

In terms of recommendations for adaptations, patient (4/11, 36%) and clinician (13/15, 86%) participants provided various suggestions. Mostly, these pertained to making Canadian-specific adaptations to content on finding work, study, and volunteer opportunities (job interviews, job-finding tools, job hunting and study searching, work, study, and life) and content on postdischarge care (finding a doctor, paying for postdischarge support, and support after discharge). They also recommended having the audio tracks read by individuals with a Canadian accent (vs Australian) and adding content pertaining to physical health (eg, adding a section on sleep hygiene), medication use and side effects, aptitude tests and self-rating scales, information on nondistressing psychosis, and links to other websites and resources:

> *Definitely agree that there needs to be more resources connected to the website. So, you know, it's not just one mental health issue, it's many issues, so it's best to know what's out there.* [P12]

In terms of the features of the site, clinicians recommended reducing the volume of text and adding more audio tracks to deliver the content, given that youths recovering from FEP can have:

> *residual cognitive difficulties which could limit their ability to read through text for extended time periods.* [C8]

Adapting the Digital Health Innovation

Subsequent to the data collection phases of this study, we proceeded to develop the content for the module adaptations, especially in relation to employment, education, and postdischarge. This process was conducted iteratively (ie, back and forth communications and reviews until a satisfactory outcome was achieved) in consultation with four of the clinician participants (2 urban and 2 urban-rural) and two of the patient participants (1 urban and 1 urban-rural). With respect to the topics of employment, study, and volunteer opportunities, modifications were made to the following sections on Horyzons: (1) job-finding tools, (2) job hunting and study searching, (3) job interviews, and (4) work, study, and life. For example, modifications were made to ensure that local organizations and government agency services, resources, and links replaced the Australian ones to reflect the local provincial and municipal job market. Moreover, some of the content was adapted to Canadian norms, for example, the *do's and don'ts* of what is included in a resume.

With respect to the topic of postdischarge, modifications were made to the entire content, and the sections were named as follows: (1) getting provincial health insurance, (2) finding a family doctor, and (3) postdischarge follow-up. Clinician participants suggested that the text in these modules be presented using flowcharts and tables, and they participated in the development and visual representation of this content. For example, for the module on postdischarge, one clinician suggested a table to organize information that provides patients with a brief description of the 3 types of postdischarge follow-up care that are available to them depending on their needs.

Once a preliminary version of the adaptations was completed, it was shared with the patient participants along with a feedback form and a brief survey. One patient participant from each site completed the feedback form and the survey, and both expressed that they perceived the job-related modules to be helpful and informative. The patient participants also provided additional recommendations at this stage, such as modifications to the sample resume in terms of using "more specific or detailed information" (P1); adding an example of *cold emailing* to potential employers; expanding the section on what the provincial employment sector "offers to help seek employment" (P11); adding "more information on what is deducted from a pay," "what might constitute diminished capabilities at a place of work," and "how to balance work, school, and life so as to not get over stressed and have a relapse" (P11).

With regard to the postdischarge modules, the patient participants had limited feedback. One of the participants

suggested including more information about transitioning from disability support to gainful employment and how this relates to health insurance. This same participant also commented on needing more information pertaining to which type of postdischarge setting (eg, family doctor vs secondary outpatient psychiatry) would be most suitable for different circumstances.

Once the content of the adaptations was finalized, a separate copy of the Horyzons platform was created for Canadian use in which the adaptations were made, herein referred to as HoryzonsCa. The platform has an inbuilt authoring system, so that those responsible for localizing content can simply log on to the system and make the changes themselves, with limited need for technical support. We then began preparing for the live pilot implementation study of HoryzonsCa. We decided to focus the pilot study on the urban site, given the issues with internet connectivity at the urban-rural site. This issue may be something for the research team to consider for future projects in community settings that have limited infrastructure in terms of internet connectivity. The preparations for the pilot study resulted in additional modifications to the following: terms of use, setting the time to match our time zone (ie, Eastern Standard Time), making the *need help now* button more apparent by changing the font color to red and adapting the content of the *need help now* page to the local context.

MOST contains a mechanism whereby client newsfeed postings containing problematic words, as contained in a system word list, are automatically blocked from being posted. This list of words consists of both swear words and danger words (ie, words indicating that the user is at risk). Postings with swear words are blocked, and the user is prompted to provide an appropriately reworded version of the post. Posts containing danger words are automatically sent to a moderation quarantine, where they are assessed by a moderator and either deemed okay and released to the newsfeed or deemed problematic, prompting a check in with the user. We translated the Australian list of words to French and made any other culturally required modifications even though the implementation was planned to be in English. This is because many of the youths seen in the program are bilingual, with their mother tongue being French. As such, it was considered that some of these individuals might naturally revert back to their mother tongue during crisis situations and moments of disinhibition. Thus, to increase the safety of the platform, we added translated French words to the trigger list. We also made changes to the moderation interface of the platform, including the contact list, safety check and supervision roster, and the structure and process for the moderators to document their notes to fit with local practices (see Table 4 for a summary of the adaptations made to the Horyzons platform).

Table 4. Summary of adaptations.

Original features	Participant feedback	Adapted version
Job-finding tools module	Keep the structure and integrate the local content	• Canadian *do's and don'ts* of writing a resume and a cover letter • A sample resume and cover letter
Job-hunting and study-searching module	Keep the structure and integrate the local content	• Description and links to provincial and local job-seeking sites, government agencies, employment centers, major employers (eg, retail and fast food), LinkedIn, recruitment agencies, student internships (for urban areas only), web-based resources and career guides, career guidance professionals, and finding volunteer work • Deleted one section on a training and university program only relevant to Australia
Job interviews module	Keep the structure and integrate the local content	• No modifications
Work, study, life module	Keep the structure and integrate the local content	• Description and link to the provincial employment guide • Provincial and local answers to the drop-down menu of questions related to pay and conditions, health, and work
Finding a doctor module	Completely adapt the content and visual display of information to make it locally relevant and easy to navigate	• New module name: Finding a family doctor • Flowchart with *yes* and *no* questions with provincial content and links
Paying for postdischarge support module	Completely adapt the content and visual display of information to make it locally relevant and easy to navigate	• New module name: Getting provincial health insurance • Flowchart with *yes* and *no* questions with provincial content and links
Support after discharge module	Completely adapt the content and visual display of information to make it locally relevant and easy to navigate	• New module name: Postdischarge follow-up • Local postdischarge follow-up options detailed in a visual format
General use and safety features	—[a]	• Adapted terms of use • Changed to local time zone • Enhanced the *need help now* button and integrated local content
Moderation features	—	• Translated Australian list of problematic words to French because local youths taking part in the pilot phase may be bilingual, with their mother tongue being French • Adapted the following to local content and practices: contact list, safety check, supervision roster, and structure and process for moderators' notes

[a]Participants did not give specific adaptation feedback on the topic.

Discussion

Principal Findings

Overall, the HoryzonsCa beta version was well received by both patient and clinician participants. In particular, they appreciated the strengths-based therapeutic approach of the platform and considered the content and social media features to be supportive of the recovery process. Their interest and positive comments about this digital health innovation provide us with the validation for moving forward with a live pilot implementation study of HoryzonsCa. Across the various stages of the adaptation study, both clinician and patient participants provided different types of comments, suggestions, and questions. Feedback offered insights into attitudes toward the innovation, potential barriers to its implementation, and adaptations to make before piloting the intervention within a Canadian context. Some of the suggestions from participants in this study are consistent with those reported in previous research on Horyzons (eg, making the intervention accessible via mobile devices) [11]; concurrently, many of the suggestions for adaptations and improvement are novel. This may be due to the qualitative nature of our process, the elicitation of feedback from multiple stakeholder groups (ie, clinicians in addition to patients), the multiple opportunities given to reflect and contribute, the active engagement of participants in the adaptation process, and the factors related to context (eg, different characteristics of the employment and health sector). In addition, many of our patient participants had sociodemographic characteristics that may have facilitated receptivity to Horyzons, for example, several had high levels of education, which has been associated with active use of the internet [20], seeking medical information on the web [21], and the depth and breadth of engagement with web-based interventions [22] among those with psychosis and other psychotic disorders.

Beyond the content-specific modifications for Canadian applicability, this adaptation study also highlighted several issues for future consideration. First, we learned of the utility of having a framework to guide the adaptation process. The digital health innovation adaptation framework [1] that we used was feasible to apply and provided a clear pathway for the adaptation process. We also learned that digital health adaptation

processes should incorporate multiple opportunities over time for users to provide feedback on how a digital health innovation can be optimized to address their needs. However, it is also important to consider that tailored adaptations of a digital health innovations across multiple settings and population groups may raise challenges for managing the technology infrastructure on the back end; this could be an area warranting consideration in future digital mental health innovation research.

We also learned from this study that a multiple stakeholder consultation approach to digital health adaptation can create challenges in terms of meeting a diverse set of expectations. We balanced this challenge by prioritizing adaptations that were raised by multiple participants (ie, patterns) and through collective reflection on which adaptations were necessary before live implementation and which suggestions for adaptations could be considered as secondary. It is also important to note that none of the suggestions that were made jeopardized the core features of the intervention (eg, peer support or clinician moderation) nor its theoretical underpinnings (eg, self-determination theory).

Second, it is important to consider the universal accessibility of digital health innovations, including, for example, accessibility to individuals with visual, cognitive, and physical impairments. The latest version of the platform has begun to address some of these issues, for example, by reducing the amount of text and conveying information through visual media, such as comic illustrations. Third, it is important to consider that adaptations to a digital health innovation must align with jurisdictional guidelines for privacy and security. Fourth, as several participants outlined the lack of access to technology infrastructure as a barrier, it is important to optimize or configure the intervention for ease of navigation and access by communities living in urban-rural, rural, and remote settings with limited internet speed.

Finally, it is important to consider workforce capacity, training, and readiness in terms of providing blended approaches to care (eg, case management in person and on the web). With respect to the latter, the concerns raised by clinicians regarding moderating the site may be linked to limited knowledge and experience with using technology-based interventions and services. This lack of experience highlights the ongoing need for continuing education and training on digital health for the mental health workforce [23]. Aligned with this, it is also important to consider whether clinical training materials to implement digital health innovation need to be adapted for use in a specific context. Furthermore, there may be a need to consider new models for moderating digital health innovations that combine the human resource capacity across a range of stakeholders (eg, case managers, peer support workers, clinician moderators).

Conclusions

With the majority of young people now using the internet, social media, and mobile devices, the use of technology is an important avenue for supporting recovery in young people experiencing FEP. Research on the use of technology to deliver psychosocial interventions is emerging within the field of early intervention for psychosis [24-27]; however, less attention has been given to how digital health innovations can be adapted across settings. Our study addresses this gap and contributes to the growing field of implementing digital health innovations within the context of mental health service delivery. Specifically, this adaptation study provides insights into the processes and outcomes of adapting and testing an existing digital health innovation for use in different contexts and health care settings. Using a systematic approach, this study identified specific aspects of a digital health intervention perceived by clinicians and patients as being important to adapt to enhance the contextual implementation and acceptability of the intervention. Our next plan is to test the acceptability, safety, and potential benefits (eg, clinical efficacy) of the adapted platform using a live version of the site with a sample of 20 to 25 participants over an 8-week follow-up period, with access to clinician and peer support moderators. Finally, in the context of the COVID-19 pandemic and public health guidelines for social distancing, there is an increasing interest and need to leverage internet and mobile technologies for delivering youth mental health services. As countries look to one another for guidance on how to navigate the changing social dynamics, knowledge on how to utilize and adapt existing innovations across cultural contexts is now more important than ever. Our study provides insights into the processes and methods that can be used to prepare for transporting a digital health innovation across settings and geographical contexts and lessons learned through this process.

Acknowledgments

The authors gratefully acknowledge the contributions of all the patient and clinician participants who dedicated their time to this study. This Phase 1—Adaptation Study of the Horyzons-Canada research program was supported by a NARSAD Young Investigator Grant from the Brain and Behaviour Research Foundation awarded to SL. During this study, SL was supported in part by a Research Scholar Salary Award from the Fonds de recherche du Québec—Santé and subsequently in part by a New Investigator Salary Award from the Canadian Institutes of Health Research. AM and SL are currently funded in part by the Canada Research Chairs Program. MA was supported by an investigator grant (APP1177235) from the National Health and Medical Research Council of Australia.

Conflicts of Interest

SL reports a recent research grant from Hoffman-La Roche, pertaining to an upcoming phase 3 study on the implementation and evaluation of HoryzonsCa; MA, JG, and RJ are co-investigators on this grant. RJ served as speaker and member of advisory

board committees for Pfizer, Janssen, BMS, Sunovian, Myelin and Associates, Otsuka, Lundbeck, Shire, and Perdue. He also received grants from Janssen, BMS, Otsuka, Lundbeck, Astra Zeneca, and HLS. All of these are unrelated to this study. AM has received funding for research consultations and honoraria for lectures delivered at conferences sponsored by Lundbeck and Otsuka, Canada, and Global, unrelated to this study.

Multimedia Appendix 1
Focus group room setup at the urban-rural site. Focus group room setup at the urban site.
[PDF File (Adobe PDF File), 249 KB - formative_v4i10e19887_app1.pdf]

Multimedia Appendix 2
Coding framework.
[PDF File (Adobe PDF File), 53 KB - formative_v4i10e19887_app2.pdf]

References

1. Lal S, Gleeson J, Malla A, Rivard L, Joober R, Chandrasena R, et al. Cultural and contextual adaptation of an ehealth intervention for youth receiving services for first-episode psychosis: adaptation framework and protocol for Horyzons-Canada phase 1. JMIR Res Protoc 2018 Apr 23;7(4):e100 [FREE Full text] [doi: 10.2196/resprot.8810] [Medline: 29685867]
2. Harper Shehadeh M, Heim E, Chowdhary N, Maercker A, Albanese E. Cultural adaptation of minimally guided interventions for common mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016 Sep 26;3(3):e44 [FREE Full text] [doi: 10.2196/mental.5776] [Medline: 27670598]
3. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995 Feb;23(1):67-82. [doi: 10.1007/BF01447045] [Medline: 7759675]
4. Castro FG, Barrera M, Holleran Steiker LK. Issues and challenges in the design of culturally adapted evidence-based interventions. Annu Rev Clin Psychol 2010;6:213-239 [FREE Full text] [doi: 10.1146/annurev-clinpsy-033109-132032] [Medline: 20192800]
5. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
6. Sundell K, Beelmann A, Hasson H, von Thiele Schwarz U. Novel programs, international adoptions, or contextual adaptations? Meta-analytical results from German and Swedish intervention research. J Clin Child Adolesc Psychol 2016;45(6):784-796. [doi: 10.1080/15374416.2015.1020540] [Medline: 25864716]
7. Chowdhary N, Jotheeswaran AT, Nadkarni A, Hollon SD, King M, Jordans MJ, et al. The methods and outcomes of cultural adaptations of psychological treatments for depressive disorders: a systematic review. Psychol Med 2014 Apr;44(6):1131-1146 [FREE Full text] [doi: 10.1017/S0033291713001785] [Medline: 23866176]
8. Smith TB, Rodríguez MD, Bernal G. Culture. J Clin Psychol 2011 Feb;67(2):166-175. [doi: 10.1002/jclp.20757] [Medline: 21105069]
9. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]
10. Gleeson JF, Lederman R, Wadley G, Bendall S, McGorry PD, Alvarez-Jimenez M. Safety and privacy outcomes from a moderated online social therapy for young people with first-episode psychosis. Psychiatr Serv 2014 Apr 1;65(4):546-550. [doi: 10.1176/appi.ps.201300078] [Medline: 24687106]
11. Ludwig KA, Browne JW, Nagendra A, Gleeson JF, D'Alfonso S, Penn DL, et al. Horyzons USA: a moderated online social intervention for first episode psychosis. Early Interv Psychiatry 2020 Feb 17:- epub ahead of print. [doi: 10.1111/eip.12947] [Medline: 32067415]
12. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104. [doi: 10.1136/bmjopen-2018-024104] [Medline: 30782893]
13. Alvarez-Jimenez M, Gleeson JF, Bendall S, Penn DL, Yung AR, Ryan RM, et al. Enhancing social functioning in young people at ultra high risk (UHR) for psychosis: a pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
14. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
15. McEnery C, Lim MH, Knowles A, Rice S, Gleeson J, Howell S, et al. Social anxiety in young people with first-episode psychosis: pilot study of the EMBRACE moderated online social intervention. Early Interv Psychiatry 2019 Dec 30:- epub ahead of print. [doi: 10.1111/eip.12912] [Medline: 31889431]

16. Bailey E, Alvarez-Jimenez M, Robinson J, D'Alfonso S, Nedeljkovic M, Davey CG, et al. An enhanced social networking intervention for young people with active suicidal ideation: safety, feasibility and acceptability outcomes. Int J Environ Res Public Health 2020 Apr 03;17(7):2435 [FREE Full text] [doi: 10.3390/ijerph17072435] [Medline: 32260111]
17. Gleeson J, Lederman R, Koval P, Wadley G, Bendall S, Cotton S, et al. Moderated online social therapy: a model for reducing stress in carers of young people diagnosed with mental health disorders. Front Psychol 2017;8:485 [FREE Full text] [doi: 10.3389/fpsyg.2017.00485] [Medline: 28421012]
18. Malla A, Lal S, Vracotas NC, Goldberg K, Joober R. Early intervention in psychosis: specialized intervention and early case identification. Encephale 2010 Mar;36(Suppl 3):S38-S45. [doi: 10.1016/S0013-7006(10)70029-4] [Medline: 21095391]
19. Implementing the Early Intervention on Psychosis Access and Waiting Time Standard: Guidance. NHS England, the National Collaborating Centre for Mental Health and the National Institute for Health and Care Excellence. 2016 Apr. URL: https://www.england.nhs.uk/mentalhealth/wp-content/uploads/sites/29/2016/04/eip-guidance.pdf [accessed 2020-10-07]
20. Välimäki M, Kuosmanen L, Hätönen H, Koivunen M, Pitkänen A, Athanasopoulou C, et al. Connectivity to computers and the Internet among patients with schizophrenia spectrum disorders: a cross-sectional study. Neuropsychiatr Dis Treat 2017;13:1201-1209 [FREE Full text] [doi: 10.2147/NDT.S130818] [Medline: 28490882]
21. Baup H, Verdoux H. Frequency and pattern of Internet use in patients with schizophrenia or bipolar disorders seeking medical information. Psychiatry Res 2017 Jan;247:152-154. [doi: 10.1016/j.psychres.2016.11.028] [Medline: 27893996]
22. Arnold C, Villagonzalo KA, Meyer D, Farhall J, Foley F, Kyrios M, et al. Predicting engagement with an online psychosocial intervention for psychosis: exploring individual- and intervention-level predictors. Internet Interv 2019 Dec;18:100266 [FREE Full text] [doi: 10.1016/j.invent.2019.100266] [Medline: 31890619]
23. Lal S. E-mental health: promising advancements in policy, research, and practice. Healthc Manage Forum 2019 Mar;32(2):56-62. [doi: 10.1177/0840470418818583] [Medline: 30739487]
24. Valentine L, McEnery C, O'Sullivan S, Gleeson J, Bendall S, Alvarez-Jimenez M. Young people's experience of a long-term social media-based intervention for first-episode psychosis: qualitative analysis. J Med Internet Res 2020 Jun 26;22(6):e17570 [FREE Full text] [doi: 10.2196/17570] [Medline: 32384056]
25. Valentine L, McEnery C, Bell I, O'Sullivan S, Pryor I, Gleeson J, et al. Blended digital and face-to-face care for first-episode psychosis treatment in young people: qualitative study. JMIR Ment Health 2020 Jul 28;7(7):e18990 [FREE Full text] [doi: 10.2196/18990] [Medline: 32720904]
26. Bucci S, Morris R, Berry K, Berry N, Haddock G, Barrowclough C, et al. Early psychosis service user views on digital technology: qualitative analysis. JMIR Ment Health 2018 Oct 31;5(4):e10091 [FREE Full text] [doi: 10.2196/10091] [Medline: 30381280]
27. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]

Abbreviations

FEP: first-episode psychosis
MOST: Moderated Web-based Social Therapy
SEI: specialized early intervention

Edited by G Eysenbach; submitted 01.07.20; peer-reviewed by B Ridout, C Arnold; comments to author 15.08.20; revised version received 31.08.20; accepted 22.09.20; published 29.10.20.

Please cite as:
Lal S, Gleeson J, Rivard L, D'Alfonso S, Joober R, Malla A, Alvarez-Jimenez M
Adaptation of a Digital Health Innovation to Prevent Relapse and Support Recovery in Youth Receiving Services for First-Episode Psychosis: Results From the Horyzons-Canada Phase 1 Study
JMIR Form Res 2020;4(10):e19887
URL: http://formative.jmir.org/2020/10/e19887/
doi:10.2196/19887
PMID:33118945

Original Paper

Consumer-Guided Development of an Engagement-Facilitation Intervention for Increasing Uptake and Adherence for Self-Guided Web-Based Mental Health Programs: Focus Groups and Online Evaluation Survey

Amelia Gulliver[1], BA, BScPsych (Hons), PhD; Alison L Calear[1], BAppPsych (Hons), PhD; Matthew Sunderland[2], BPsych (Hons), PhD; Frances Kay-Lambkin[3], BSc (Hons), PhD; Louise M Farrer[1], BPsych (Hons), PhD; Michelle Banfield[1], BSc, BA (Hons), PhD; Philip J Batterham[1], BSc (Hons), MPH, PhD

[1]Centre for Mental Health Research, Research School of Population Health, The Australian National University, Canberra, Australia

[2]The Matilda Centre for Research in Mental Health and Substance Use, University of Sydney, Sydney, Australia

[3]Priority Research Centre for Brain and Mental Health, University of Newcastle, Newcastle, Australia

Corresponding Author:
Amelia Gulliver, BA, BScPsych (Hons), PhD
Centre for Mental Health Research
Research School of Population Health
The Australian National University
Acton
Canberra, 2601
Australia
Phone: 61 26125 ext 9472
Email: amelia.gulliver@anu.edu.au

Abstract

Background: Self-guided web-based mental health programs are effective in treating and preventing mental health problems. However, current engagement with these programs in the community is suboptimal, and there is limited evidence indicating how to increase the use of existing evidence-based programs.

Objective: This study aims to investigate the views of people with lived experience of depression and anxiety on factors influencing their engagement with self-guided web-based mental health (e–mental health) programs and to use these perspectives to develop an engagement-facilitation intervention (EFI) to increase engagement (defined as both uptake and adherence) with these programs.

Methods: A total of 24 community members (female=21; male=3) with lived experience of depression and anxiety or depression or anxiety alone participated in 1 of 4 focus groups discussing the factors influencing their engagement with self-guided e–mental health programs and the appearance, delivery mode, and functionality of content for the proposed EFI. A subsequent evaluation survey of the focus group participants (n=14) was conducted to evaluate the resultant draft EFI. Data were thematically analyzed using both inductive and deductive qualitative methods.

Results: Participants suggested that the critical component of an EFI was information that would challenge personal barriers to engagement, including receiving personalized symptom feedback, information regarding the program's content or effectiveness and data security, and normalization of using e–mental health programs (eg, testimonials). Reminders, rewards, feedback about progress, and coaching were all mentioned as facilitating adherence.

Conclusions: EFIs have the potential to improve community uptake of e–mental health programs. They should focus on providing information on the content and effectiveness of e–mental health programs and normalizing their use. Given that the sample comprised predominantly young females, this study may not be generalizable to other population groups. There is a strong value in involving people with a lived experience in the design and development of EFIs to maximize their effectiveness.

(JMIR Form Res 2020;4(10):e22528) doi:10.2196/22528

KEYWORDS

mental health; internet; anxiety; depression; technology; treatment adherence and compliance

Introduction

Common mental disorders such as depression and generalized anxiety disorders are experienced by 5% to 10% of the population each year [1-3]. Depression and anxiety can cause high levels of disability and burden [4,5]; however, only one-third of those experiencing a disorder seek professional help [6]. Mental health programs delivered on the web (*e–mental health programs*) have been proposed as a lower-cost alternative to face-to-face therapy [7]. Web-based programs may be particularly appropriate for those identified as at risk for mental health problems or those with mild-to-moderate symptoms [8]. These programs are evidence-based, often as effective as face-to-face therapy, and have the potential to lessen the impact of many of the key barriers to seeking professional help, including cost, stigma, and accessibility [9].

Barriers to Uptake and Adherence for eMental Health Programs

Despite addressing some of these critical barriers, community uptake of e–mental health programs is low [10,11]. Studies based on primary care have reported rates of uptake between 3% and 25% [12]. Factors such as awareness of e–mental health programs and community views on the effectiveness of web-based versus face-to-face therapy could be fueling this lack of uptake [13]. Accordingly, research has demonstrated a preference in the community for face-to-face therapy over e–mental health programs [11,14,15]. However, studies have also shown that most people are willing and very few people would refuse to try an e–mental health program [16,17]. Thus, it is important to determine the factors that could explain the lack of uptake in community settings. Overcoming these implementation barriers is critical to gaining the maximum benefit of e–mental health programs in the community. Public campaigns have previously been used to raise awareness of e–mental health programs [18]. However, there is a paucity of evidence regarding the effectiveness of these campaigns and other effective strategies for increasing uptake of e–mental health programs.

Reported barriers to the uptake of self-guided e–mental health programs include a lack of general education, cost of hardware and internet access, and time demands [19]. Other potentially modifiable barriers include acceptability of web-based programs and knowledge of how to use technology [19]. Further research has investigated the issue of *acceptability* barriers to e–mental health programs [12]. The acceptability of these programs is thought to be impacted by a wide range of attitudes, including concerns about data security, anxiety about the internet in general, belief that the programs will not work, poor attitudes to help-seeking in general, a general lack of knowledge about web-based programs, or concern that the programs are not endorsed by health care authorities [9,12,14,19-22].

Poor adherence is also a common feature of self-help e–mental health programs. Only about 56% of users completed their assigned web-based program in trial settings compared with 85% in face-to-face settings and 65% in guided internet-based cognitive behavioral therapy for depression [19]. Adherence is even more problematic in naturalistic settings [23]. One study indicated that only 1 in 3 community users completed a minimum of one module with only 10% of the users completing at least two modules out of 6 [24]. Reasons for poor adherence are complex as the drop outs may be because of the lack of need (eg, healthy users), symptom remission, lack of response to treatment, lack of engaging or relevant content, or a high level of symptoms that can interfere directly with the ability to engage (eg, through low motivation or fatigue) [23-26]. Together, these factors act as barriers to the widespread adoption of potentially effective e–mental health programs, limiting the ability of these technologies to deliver on their potential.

Engagement-Facilitation Interventions

This study adopts a model of engagement [27,28] that includes both the initiation of the program (uptake) and its continued use (adherence). Research on the theory of planned behavior [29] suggests that if the factors affecting both uptake and adherence in e–mental health programs are addressed, we might be able to increase the overall *engagement* with these programs. Acceptance-facilitation interventions (AFIs) have been described as brief programs to increase acceptance of e–mental health programs [12]. Using the theory of planned behavior, the goal of AFIs is to alter subjective societal norms surrounding the knowledge and use of e–mental health programs [12,30]. Engagement-facilitation interventions (EFIs) are related to AFIs. AFIs aim to increase *acceptability* of internet interventions among end users [12,31]. In contrast, EFIs aim to increase *engagement*, which incorporates both uptake and adherence, by addressing factors associated with the acceptability of internet interventions and additional barriers to engagement, such as a lack of time or perceptions that the benefits of the program are not worth the investment [26]. One study found that the acceptability of e–mental health programs for depression in a primary care setting increased after watching a video-based EFI [12]. However, another study using a video-based EFI for increasing engagement with a chronic pain intervention did not demonstrate increased engagement (uptake or adherence) [32]. It was argued that this failure to find an effect may have been because of the sample's overall high level of motivation before the intervention [32]. Thus, there remains limited evidence on the effectiveness of EFIs in increasing engagement with e–mental health programs in the community. In addition, although previous studies examining barriers to engagement exist [16,26,33], very few qualitative studies with consumer groups have been conducted, and there is a paucity of research investigating the factors that facilitate the use of self-help e–mental health programs for common mental disorders.

Consumer Involvement

There is widespread acknowledgment of the importance of partnership in health and medical services with consumers who are defined as people with lived experience of the health condition of interest [34]. Effective involvement in research

can be described as that which is appropriately chosen for the task and the skills of both the *involvers*, in this case, the researchers, and the *involvees*, the consumers [35]. It is noted that, at a minimum, consultation with the target population in the creation of a service is critical as it allows for tailoring and an assessment of the appropriateness of the content, which can also improve the uptake of and engagement with services [36].

This study brings together the latest evidence in e–mental health program development and implementation and investigation of the primary factors influencing the engagement of self-guided e–mental health programs for consumers living in the community. We define engagement as addressing both the uptake and adherence to these programs. This study presents the development and preliminary evaluation of an EFI based on the results of a series of consumer focus groups. The resultant EFI will be tested in a randomized controlled trial [37] to assess its effectiveness in improving engagement with an existing e–mental health program.

Objectives

The aim of this study is to investigate factors influencing engagement (uptake and adherence) in e–mental health programs in a community-based sample of those with lived experience of depression and anxiety. The aim of generating this material is to inform the development of a brief EFI for a specific e–mental health program. The principles identified in the development process may be used to guide the development of a variety of future EFIs to maximize uptake and impact.

Methods

This study adheres to the Standards for Reporting Qualitative Research reporting guidelines [38]. These are presented in Multimedia Appendix 1.

Ethical Approval

Approval for the ethical conduct of the study was granted by The Australian National University Human Research Ethics Committee (ANU HREC 2018/257).

Participants

Table 1 presents the demographic characteristics of those who participated in the focus groups and the evaluation survey. The focus group participants were 24 community members aged between 18 and 70 years, who were identified as having lived experience of depression and anxiety or depression or anxiety alone. Approximately 60% (14/23) of these members participated in the evaluation survey.

Table 1. Participant demographic information for the focus groups and subsequent evaluation survey.

Demographic data	Focus groups (n=24)	Evaluation survey[a] (n=14)
Age (years)		
Mean (SD)	27.9 (12.4)	31.6 (15.3)
Range	18-70	18-70
Gender, n (%)		
Female	21 (87)	14 (100)
Male	3 (12)	0 (0)
Study status, n (%)		
University student	15 (62)	9 (64)
Nonstudent	9 (37)	5 (36)

[a]The survey participants were a subset of the focus group participants.

Recruitment

We recruited participants through advertisements posted in various local community and university Facebook groups and via direct email to local consumer and caregiver groups in the Australian Capital Territory. We specifically sought to sample a group of relevant consumers who would be a natural target audience for an e–mental health program; this included targeting a sample of young people via a university Facebook group.

Inclusion Criteria

To ensure that the group was relevant to the development of the EFI, inclusion criteria listed the characteristics of our targeted e–mental health program users. Participants were required to be adults (aged 18 years or above) who self-identified as having a lived experience of mild-to-moderate depression or anxiety, with no severe distress or suicide plan at present. Given the small number of participants, minimal demographic information was collected to reduce the risk of participants being identified.

Procedure

Focus Groups

We conducted 4 focus groups in July 2018 in Canberra, Australia. A total of 24 adults participated across the 4 focus groups (n=5, 8, 6, and 5), which ran for approximately 1-hour each. The groups were moderated by author AG (a lived experience researcher) and an assistant (NK, a research assistant or LF, a mental health clinician). All participants provided written consent to participate in the focus groups and completed a brief demographic survey. Focus group participants were advised of the respectful, voluntary, and confidential nature of the discussion and its purpose. We conducted the groups according to the principles of participatory design and iterative

development, whereby the potential users of a product or service are involved in its design [36]. Focus group discussions were digitally recorded and transcribed by a professional transcription service. A research assistant or clinician (NK or LF) recorded field notes during the session. Gift cards of Aus $50 (US $35) were distributed to thank participants for their involvement before the discussion sessions.

Evaluation Survey

After the completion of the focus groups, the points discussed were used to inform the development of a draft EFI. As a separate stage of the research, we sent an email to focus group participants inviting them to participate in a single focus group to evaluate the resultant draft EFI. The demand for this final group was unexpectedly high with 21 out of 24 participants (21/24, 88%) expressing an interest. To maximize the data collected from this final stage of the research, a web-based survey was offered to collect the participants' views. Out of the 21 interested participants who were sent a link to complete the survey, 14 (67%) responded. They were sent an e–gift card of Aus $35 (US $25) to thank them for their participation.

Data Collection

Focus Groups

Focus groups were selected as they can capitalize on the interaction among participants to generate richer data than that obtained from individual interviews [39]. However, a limitation of this method is that depending on group dynamics, some participants may find it difficult to voice their opinion. Thus, a modified nominal group technique [40] was used to ensure that all participants' ideas could be heard. A phase of *silent idea generation* [41] provided all participants with an opportunity to provide ideas for what factors may increase or decrease their engagement in e–mental health programs. This was followed by a group discussion and then an individual ranking activity to determine the relative importance of barriers to engagement, which were previously discussed.

Multimedia Appendix 2 presents the list of focus group questions for each of the 4 groups (groups 1-4). The discussions were highly structured and based on predetermined topics sourced by the authors from previous reviews of the literature [9,12,14,19-21]. Participants were initially provided with a verbal description and visual examples of self-guided e–mental health programs to ensure clear understanding of the topic before the discussion. Overall, 2 key topics were discussed in each group, such as (1) factors decreasing (barriers) and increasing (facilitators) their potential engagement in web-based self-guided mental health programs and (2) preferences for EFI content and presentation (eg, video, text, audio, presenters).

Multimedia Appendix 3 presents the written activities that were also completed by participants within the focus groups. The activities were designed to identify the most critical factors decreasing (barriers) and increasing (facilitators) their engagement in e–mental health programs. Participants were instructed to write, on their worksheet, 3 things that might stop them from engaging in a web-based self-guided mental health program (activity 1, targeting both uptake and adherence) and 3 things that could help keep them engaged in a web-based self-guided mental health program (activity 3, primarily targeting adherence). They were also asked to rank their top 3 barriers to engaging in web-based self-guided mental health programs by numbering them (1, 2, and 3) in the order of the barriers they felt were the most important in stopping them from engaging with these programs (activity 2, targeting both uptake and adherence). Participants in focus groups 2 to 4 were also presented with a basic EFI prototype to comment on, which was iteratively developed based on the discussion and input of previous focus groups.

Relevant probing questions were used to gather further details where necessary. The researcher (AG) conducting the focus groups used their lived experience to develop rapport with the participants and to encourage open responses. The groups were continued until the lead author was satisfied that theoretical saturation of ideas about the factors influencing engagement and design concepts for the EFI had occurred (no new insights generated from the data) [42].

Evaluation Survey

Multimedia Appendix 4 presents the questions included in the evaluation survey to gauge focus group participants' opinions on the draft EFI developed as a result of the focus groups. The survey consisted of a range of quantitative and qualitative measures assessing what the participants liked and disliked about the EFI, what was missing from the EFI, the likely effect the EFI would have on their decision to start and complete an e–mental health program, their satisfaction with the EFI, and identifying to whom the EFI may appeal.

Analytical Strategy

Thematic analysis [43] was used by the first author (AG) to group key concepts into themes. For the ranking activity as part of the modified nominal group technique, the relative importance of the ranked barriers was assessed by reverse scoring the ranks, which were then cumulated across participants and tabulated. Thematic analysis of the focus group discussion on both factors influencing engagement and the proposed presentation of the EFI was primarily deductive. Structured questions that were developed ahead of the focus groups based on previous literature on engagement in e–mental health programs were used to develop themes for key factors for inclusion in the EFI. This was complemented by themes generated inductively from written activities that were used to create thematic concept maps. Data from the written activities are presented first, followed by the focus group discussion on factors influencing engagement, including the proposed presentation of the EFI. Finally, the results of the evaluation survey that evaluated the draft EFI are presented. To protect privacy, participants were identified using a participant number.

Results

Focus Group

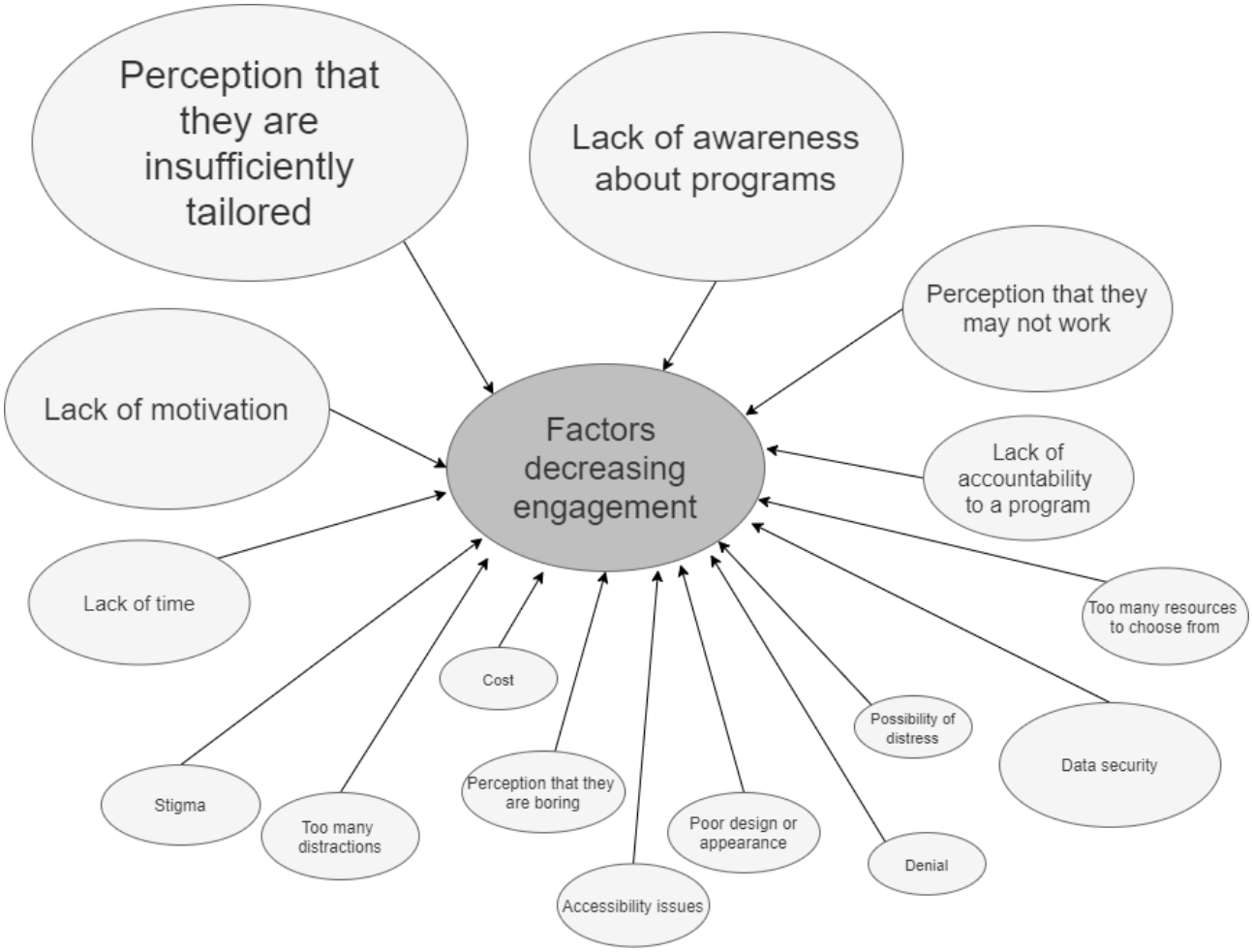
Written Activities

Factors Decreasing Engagement

Figure 1 presents a concept map of the factors that participants identified as likely to decrease their engagement with self-guided

e–mental health programs from written activities. The most common factors or barriers to engagement identified were (1) perception that the programs could not be sufficiently tailored or personalized because of their web-based delivery (15/74, 20%), (2) perception of a lack of awareness about the existence of web-based programs (14/74, 19%), (3) perception of difficulties with self-motivation to complete a web-based program (11/74, 15%), and (4) perception that e–mental health programs may not work (9/74, 12%).

Figure 1. Factors proposed to decrease engagement from the written activity. Larger shapes indicate a greater number of factors reported in that theme.



Ranked Barriers to Engagement

Table 2 provides a summary of the ranked barriers to engaging with an e–mental health program. Overwhelmingly, the most highly ranked factor was related to not knowing whether the program would help them. This was followed by a general lack of awareness of e–mental health programs.

Table 2. Barriers to engaging in e–mental health programs as ranked by participants in the written activity.

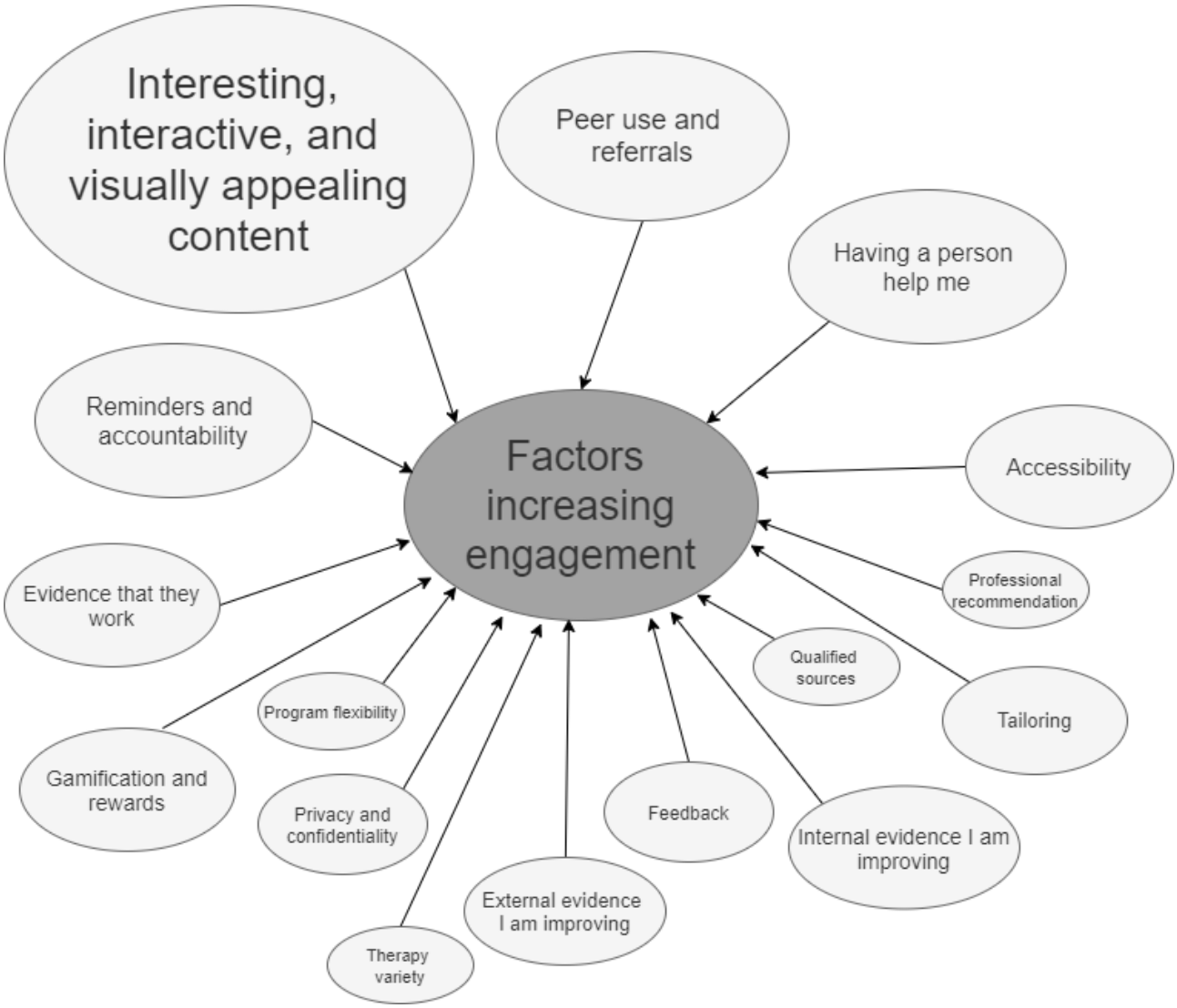
Rank	Barriers	Score[a]
1	"I don't know if the online program will help me"	48
2	"I'm not aware of online mental health programs"	19
3	"The online program is too hard to use"	10
3	"I think that I should be able to solve my problems on my own"	10
5	"I'm worried about data security"	9
6	"I'm not comfortable or familiar with existing online mental health programs"	6
6	"I'd be worried about someone finding out I was using an online mental health program"	6
8	"I'm worried that using online mental health programs isn't something normal, that lots of people do"	3
8	"I don't think any of these would stop me engaging in online self-guided mental health programs"	3
8	"Other (please explain)—User interface design"	3
8	"Other (please explain)—Not understanding or relating to the material, or activity, and knowing if you did it correctly"	3
8	"Other (please explain)—Less accountability to engage than face to face"	3
8	"Other (please explain)—Too many options (overwhelming)"	3
14	"Other (please explain)—They do not present an accurate picture of my mental health"	2
14	"Other (please explain)—No feedback/validation"	2
14	"Other (please explain)—The amount of effort to use it"	2
14	"Other (please explain)—Feeling that face-to-face mental health help is more useful"	2
14	"Other (please explain)—I feel I need a therapist"	2
19	"Other (please explain)—Don't know what the program entails/what to expect"	1
19	"Other (please explain)—They make me feel more depressed because it reinforces that there is no one there who cares (wants to talk to me)"	1
19	"Other (please explain)—Lacking the motivation to access the program"	1
19	"Other (please explain)—Expecting that there will be a lot of effort involved"	1
19	"Other (please explain)—I do not want to put in the effort to a potentially dull process"	1
24	"I feel anxious about using the internet overall"	0

[a]Ranks are reverse scored and cumulated across participants. Scores for each topic were calculated by cumulating the reverse-scored ranks (ie, 1=3, 2=2, and 3=1) across participants. Higher scores indicate higher importance.

Factors Increasing Engagement

Figure 2 presents a visual depiction of the factors considered to increase participants' engagement with e–mental health programs from the written activities. The most common factors or facilitators of engagement included (1) a need for the content to be *interesting and engaging* (15/71, 21%) and (2) the knowledge that *peers had used the program* (9/71, 13%). Despite being specifically asked about self-guided programs, several participants (8/71, 11%) indicated a need for outside input from a mental health professional such as a psychologist. A smaller number of comments suggested that the program needed to be easily accessible (eg, works on phones; 5/71, 7%), and some comments expressed a desire for gamification and rewards (4/71, 6%).

Figure 2. Factors proposed to increase engagement from the written activity. Larger shapes indicate a greater number of factors reported in that theme.



Focus Group Discussion

EFI Development

Overall, participants identified 2 main concerns that needed to be addressed in an EFI to increase user engagement, including both uptake and adherence to a self-guided e–mental health intervention. These were information about (1) what the program involved and what they could expect (in terms of activities, time commitments, etc) and (2) proof of the program's effectiveness from both scientific (information) and peer (*testimonial*) sources. Furthermore, desired components included feedback about symptoms and information about privacy and confidentiality. Participants also provided suggestions on the style and presentation of the overall EFI, noting a strong preference for a simple linear click-through format rather than having the content presented on a single page. Participants are referred to by numbers (ie, participant number 1=P1).

Key Factors for Inclusion in an EFI

Awareness

Most participants were not aware of the e–mental health programs that we presented as examples in the focus groups. This lack of awareness was noted by participants as a key factor that reduced e–mental health program uptake in the community. Although e–mental health programs were considered beneficial for help-seeking, particularly for those who felt uncomfortable with face-to-face care, participants felt, "you have to know of its existence [first]" (P1).

Information About What the Program Involves

Overall, participants reported a strong need for an EFI to contain information about what would be involved in completing the subsequent e–mental health program. This included descriptions of the therapy type and the activities they would be required to complete. One participant remarked that they wanted to know about the specific therapy types (eg, cognitive behavioral therapy). Before participating in a program, participants also wanted to know how structured or prescribed the overall program would be, the frequency and duration of recommended usage, and also the number of sessions or timeframe before they could expect to feel better. One participant thought that it would be motivating to display the potential for symptom improvement on a graph—"here's where you are and with use of our program we predict you could be all the way down to here" (P2).

Evidence of the Program's Effectiveness

To increase the likelihood of starting and continuing to use a program, participants also reported the need of evidence of a program's effectiveness based on both scientific sources and peer reviews or testimonials. Many participants reported a desire for evidence in the EFI that a web-based program works for people "like them" or with similar psychological problems. They also wanted to know from the EFI that the e–mental health program was tailored and could adapt to each person, that it could handle more than one problem at once (comorbidity), and that it was worth the time investment:

> *When you're feeling a bit low or something it takes extra effort to go in and do something and if you're going to spend a lot of your limited energy on something and you don't think there's going to be of a difference from where you are now then you might not bother with it if it's not very effective.* [P3]

It was considered particularly important for increasing uptake and adherence for an EFI to demonstrate the effectiveness of web-based programs specifically, as participants noted that there remains, in the community, strong "perceptions that [e–mental health programs are] not as effective as face-to-face" (P4).

Participants also suggested that the EFI could include evidence of effectiveness from peers who had used the program. It was a common view expressed across all 4 groups that multiple and varied genuine testimonials (*reviews*) were desired as further *proof* that the program worked. Viewing a review from someone they trust who has had success with the program was considered as, or even more, important than evidence from scientific or professional sources. However, these reviews had to appear genuine, as participants believed that they could tell if they were faked—usually they were viewed as disingenuous if the reviews were too positive or if there were no negative reviews. In addition, if the people in the reviews were too attractive or not sufficiently varied in appearance, these were also viewed as simulated:

> *There needs to be a variety. I think you can't just have textbook 30 something woman, you know, aesthetic face. I think there needs to be a variety of ages because I think these things need to cater to older generations as well as younger ones. And also a variety of ethnicities I think is very important…They can't all be beautiful.* [P5]

The participants also wanted reviews to take into account complex and relevant issues such as outlining what is different, better, or worse than face-to-face therapy.

Other Concepts

Several other topics were also discussed to a lesser degree, including the provision of feedback, privacy, and confidentiality information, emergency contacts, and cost. Some participants believed that it was important to provide feedback on symptoms, particularly before starting a web-based self-guided mental health program to help them accept that they needed help. They also thought that it would be helpful if the EFI could provide a description of what type of mental health problem they likely had and evidence that a program could be effective for and tailored to that particular type of mental health problem. Some participants wanted financial information to be included in the EFI about how much it would cost them personally to use the program, as this was considered as important information in their decision to start the program. Other information desired was how privacy and confidentiality would be maintained. Some believed privacy was the most important factor for deciding whether to engage in these programs; whereas, many others felt that it was now part of modern day life to give up some privacy to access web-based or mobile services and that "this is just the way the world is now, everyone knows everything...my phone's listening to me right now" (P6).

However, this did not extend to all information; there was a preference for providing less personal information:

> *It doesn't worry me that much. Particularly 'cause I'm not putting in my super, super personal information. Yeah, if I was...write down my address and you know the names of these people that you've seen and your GP, and whatever, maybe that would worry me more.* [P7]

Conversely, participants with concerns relating to their place of employment shared that they may not start an e–mental health program without evidence that the program was completely confidential and that their data would be kept private.

Optimal Presentation for the EFI

Participants also provided advice about the overall style and presentation of the EFI, as this would also impact their likelihood of engaging in a program.

Overall Presentation and Delivery

In general, participants did not like the content of the EFI being displayed statically on one page; they found it overwhelming and confusing. They commented that they would prefer to have information presented slowly and step-by-step, citing that when experiencing symptoms of anxiety or depression, too much content or words on a page could be stressful. In addition, for similar reasons, all but one participant also disliked the idea of one page with the inclusion of drop-down menus to access content.

Presentation of Symptom Feedback

Participants were shown multiple versions for how feedback on symptoms could be presented in an EFI after the completion of a brief assessment of depression or anxiety: a traditional histogram-style graph, a traffic light, or a simple meter pointing to the symptom severity. In general, participants preferred the simplicity of a meter with 3 or 4 points to display the severity of symptoms. However, they did say that a small amount of explanatory text was essential so that they could understand what the feedback meant. However, some of the participants in one group did not like cutoffs for the meter, preferring a *spectrum* model of symptoms:

> *I like things that kind of have kind of a spectrum, which I think both of these do. Because…if you draw a line then that is really stressful for people who are on one side or people who are kind of getting towards that line.* [P6]

Some participants preferred the more complicated histogram-style graph, although they explicitly noted that they did not want to be presented with normative population data. They noted that they already knew how bad they felt and comparisons with other people would likely make them feel worse, particularly if their symptoms appeared more or less severe than the population or were incongruent with how they felt:

> *Yeah, I don't really want to know everyone else's because I'm sure everyone else was probably a lot better than mine and that makes me feel really ostracised…I really am depressed and everyone else is so far ahead so what's the point in trying? It would just make me feel worse* [P8]

Presentation of Information

Overall, text was the dominant preference for the display of information in an EFI; however, it was clear that this text had to be brief and preferably with bullet points for ease of reading. Participants believed that video presentation in an EFI may be preferred by some users, although, again, they indicated that it must be brief (ie, 30 seconds or less). They also said that the videos should be accessible, including subtitles or a written transcript. Suggestions for presenters within an EFI video were very broad—some participants wanted someone famous who could elicit trust, such as a well-known athlete, some wanted experts, and some mentioned that they would prefer animations or cartoons that transcend specific population groups. A strong preference for variety in the way that information was presented in an EFI emerged, and several participants mentioned that a mixture of presentations would likely help to keep them engaged. Variety was also preferred for flexibility of use, such as using audio with headphones or preferring visual data on public transport. When asked if they wanted a helper or a *guide* character to accompany them through the EFI, some participants agreed that this could be helpful, but they explicitly noted that it must be optional as "you want to encourage people to look at it and click on it, you don't want it just in their face" (P3).

Presentation for Testimonials and Reviews

When participants were asked about their preferred delivery mode for the testimonials they desired in the EFI, there was a strong preference for video as it was easier to judge whether they thought that it was real or not, and it was more believable and likely to be genuine because of the higher degree of effort involved compared with written forms. Overall, many participants stated that they wanted to see as many different testimonials as possible. This included testimonials from different genders, ages, and ethnicities so that the participants could find the one they identify themselves with as "if they're a completely different person to me, I'm just not sure if I could relate to them and their experiences" (P9).

Evaluation Survey

The draft EFI shown to the participants in the evaluation survey is presented in Multimedia Appendix 5. Table 3 presents the quantitative data collected in the evaluation survey assessing the focus group participants' views on the resultant draft EFI. Overall, participants reported that the draft EFI would likely have a modest effect on their uptake and engagement with e–mental health (*myCompass*, developed by Black Dog Institute). Participants were also satisfied with the EFI and the way it captured their suggestions, and they were satisfied to highly satisfied with their participation in the study.

Table 3. Participant satisfaction of the engagement-facilitation intervention created via the focus groups.

Question item	Participants, n	Score, mean (SD)
"What effect would the EFI[a] have on your decision to start using myCompass?"	14	4.07 (0.83)[b]
"What effect would the EFI have on your decision to complete myCompass?"	14	3.93 (0.48)[b]
"How satisfied are you with the EFI overall?"	13	4.08 (0.64)[c]
"How satisfied are you with the way the EFI has captured your suggestions?"	13	4.08 (0.49)[c]
"How satisfied are you with your participation in this study?"	13	4.54 (0.52)[c]

[a]EFI: engagement-facilitation intervention.

[b]Scored as 5=much more likely, 4=a little more likely, 3=no change, 2=a little less likely, and 1=much less likely.

[c]Scored as 5=highly satisfied, 4=satisfied, 3=neutral, 2=dissatisfied, and 1=very dissatisfied.

Participants' Views on the EFI

Participants reported approving of a number of aspects of the draft EFI, including the brief, easy-to-read information that effectively used bullet points and headings to aid understanding. Participants also reported enjoying the video testimonials but requested that there should be a greater variety and number of them. Half (n=7) of the participants praised the design, graphics, and color scheme as being "engaging but doesn't distract from the information provided" (P10), whereas 2 participants mentioned that they did not like the color scheme—the orange in particular. Finally, several participants liked the graphical feedback and the impression that it was professional, tailored to the user, flexible, and provided recommendations.

Participants were also invited to report what aspects of the EFI they did not like. A number of participants (n=5) had no comments for this question—"I like it as it is. I was actually feeling ready to get started!" (P11), whereas others reported very specific suggestions about adding clearer information about the potential cost (or the lack thereof) and the removal of a line at the end of the feedback for depression and anxiety that sought rhetorical confirmation from the participant (eg, "Does this sound right to you?"). The participant noted that this question

sounded tentative and unlikely to provide a sense of confidence in the program. These suggestions were used to change the final version of the EFI for the RCT.

Potential Influence of the EFI

Several participants noted that the EFI was likely to influence them to start the e–mental health program, with one participant indicating that "the statement that it could be as useful as medication would be very persuasive for me" (P9). However, participants disagreed on whether the time required to use the program seemed too long. Some said that the recommended timeframes, for example, 60 min to 90 min, and 15 min per module seemed onerous. Others thought that a 15-min block of time to complete a module was easily digestible. Another participant acknowledged that although the time period seemed intimidating as a commitment, they believed that the addition of the word *recommended* helped them to overcome this barrier to uptake. Several participants noted that they found the EFI presentation and content simple to read and made the e–mental health program seem achievable. Another participant also mentioned that the EFI was not overwhelming and that this was critical to them to be able to do a program while experiencing mental health problems:

> *I wanted to start using myCompass now. I was really surprised, because I dislike online therapy programs. The EFI makes it…a fun activity rather than therapy, and one which would appeal to me when I was depressed, rather than being hard work. When I am depressed it is difficult for me to do any activities, so anything that appeals to me makes it easier to do. The EFI makes it the opposite to overwhelming. When I am depressed, I am overwhelmed.* [P12]

Finally, one participant noted that the program was unlikely to encourage her to use it, given her preexisting privacy concerns about using such a program.

Groups to Which the EFI Would Appeal

A number of participants reported that the EFI looked as though it was targeted at younger people or adults (n=7), possibly because of the bright colors used and the fact that it was delivered via the internet. However, several other participants noted that it did not seem to be targeting any specific demographic, including age and gender, which they found appealing.

Discussion

Principal Findings

This study presents a detailed examination of consumers' views of factors influencing their potential engagement with self-guided e–mental health programs for depression and anxiety. In addition, it provides consumer perspectives on the development of an EFI to improve engagement with these programs. Overall, this study indicates that presenting multiple forms of evidence that an e–mental health program can be effective was seen as the primary driver of whether an EFI could improve the uptake and engagement of the e–mental health program. A critical factor preventing participants' uptake of e–mental health programs was uncertainty about whether it could help them and whether it is worth investing their time and energy. This is consistent with previous research identifying low expectations of effectiveness as an important barrier for the uptake of e–mental health programs [12]. The information needed to confirm this varied, including demonstrating that it was sufficiently tailored, showing that other people used it via testimonials or peer use, and evidence for the effectiveness of web-based programs, which have been found to be important factors for the uptake of e–mental health programs previously [10,21]. In particular, participants reported a strong desire to receive information about program effectiveness from trusted peers. This is congruent with the theory of planned behavior, noting the importance of subjective norms in changing an individual's actions [29]. It was noted as important that these people were believable and real. However, consistent with previous research, participants also wanted scientific evidence in addition to evidence presented by peers [16]. It is likely that this issue can be addressed in part by the provision in an EFI of information about effectiveness, what it involves, what they can expect by participating, and testimonials. The participants' views on videos from this study were consistent with previous research [44] where videos had been found to be the most effective form of providing testimonials. Testimonials can contain powerful messages by using identification with the person telling the story and the viewer can relate and empathize with the person [44]. Testimonials have also been found to increase belief and message uptake more than the presentation of statistical data [45].

Privacy was not considered an important factor by most participants in influencing e–mental health program usage in this study. This differs from previous studies that had found that security and confidentiality of personal information concerns are considered critical for using e–mental health [16,22]. However, certain people in the community may have a greater desire for privacy and confidentiality assurances, which highlights the importance of clearly providing this information to ensure that participants are not opting out based on its absence. Concerns about privacy are also likely to depend on how much information is sought within a digital intervention; anonymous programs are likely to cause less concern than a program requiring names, email addresses, or phone contacts.

The finding that participants did not want to see their scores displayed in the EFI relative to normative population data was unexpected, particularly because the presentation of normative data is commonly used to motivate help-seeking for multiple mental health problems, including alcohol use [46] and depression [47]. The idea behind presenting normative data is to reduce normative misperceptions about alcohol or other drug use [48] and to motivate the person to seek help. However, it is possible that this does not translate directly to other mental health problems and that there may be sensitivities and stigma associated with labeling an individual as having a certain level of symptoms. An approach taken in the alcohol or other drug usage field when presenting normative feedback may be important in this context, that is, permission is sought from the participant before providing normative feedback, giving the person some control over the type of feedback to which they

are exposed. Some participants did find the concept of graphically presenting their baseline score and their capacity to improve with treatment as important, suggesting that if feedback is provided, it may need to be in the context of messaging how (and by how much) symptoms can be improved by engaging with the e–mental health program. Potentially, an algorithm for future EFIs could be presented to people about to embark on an e–mental health program that uses average improvement scores drawn from previous research with that program to estimate the percentage symptom improvement a participant could expect based on their baseline score. However, it is important to note that it is difficult to convey the complexity of changes in severity over time using a simple illustration, so there may be unintended consequences of presenting this information, such as a sense of *failure* if expected gains are not achieved. More research on this particular issue is warranted.

There was a significant lack of awareness of available e–mental health programs among participants. Given that recruitment specifically noted the topic, this was an unexpected finding. Despite local public campaigns about e–mental health programs [18], it appears that there may be a lack of knowledge in the community about these programs. As noted by both Ebert et al [12] and Gun [21], an alternative avenue of increasing knowledge in the community about the availability and effectiveness of e–mental health programs is to better inform clinicians about e–mental health programs, as they are important gatekeepers to the usage of these programs [49]. However, engaging clinicians may be a complex process, and most people with mental disorders do not engage with mental health professionals [6]; thus, direct-to-consumer pathways are also needed to maximize the impact of e–mental health programs in the community [9,50].

Implications for Intervention Development

Consumer views are critical in the development of new interventions and resources [36]. However, it is important to recognize that perspectives often vary broadly and it is often not possible to reach group consensus [51]. In the creation of an intervention, it is important to balance the selection of components using information that combines both the best available evidence (eg, content that aligns with theory, previous research, consumer views) with the practical constraints of the intervention development. For example, it was not feasible for the EFI that was developed in this study within our budget to create multiple testimonial videos of a wide variety of genders, ethnicities, and ages. However, on balance, this factor was also not deemed to be as critical or as practical to implement as education about how e–mental health can help people in the community. Thus, a pragmatic approach to the development of interventions using consumer perspectives is recommended. Broadly, evidence from this study suggests that EFIs should include information about the purpose of e–mental health programs and its components and evidence demonstrating that the program is effective. Use of video testimonials, simple text and images, and simple feedback about symptoms are also likely to be useful.

Increasing consumer engagement with internet interventions is challenging. However, ensuring that the content is interesting, interactive, and visually appealing may increase engagement—color selection and the inclusion of graphics is particularly important. Despite explicitly examining self-guided interventions, several participants noted the need for accountability to *someone*. Having a virtual guide assist the users through a program may address the barrier of a lack of motivation to engage by addressing the perceived lack of accountability to a computerized program (as opposed to a therapist). Dynamic engagement processes through the course of an intervention, such as the use of guides and reminders, may be more useful for promoting adherence than passive information provided at the start of the intervention.

Knowing peers who had used the program or providing a referral was seen as helpful. Testimonials may provide a proxy for personal referral; however, they were seen by some to be *fake* or less compelling. This issue could potentially be addressed by the inclusion of authentic video testimonials that explicitly address potential challenges with the program, particularly by avoiding the use of actors. Ensuring that the content appears sufficiently tailored to the person is also important. Although it is difficult to ensure that all mental health problems are covered, simple tailoring strategies, such as choice of a relatable avatar, may provide a sense of agency. Finally, to use the programs, the participants said that they had to know that such programs existed. The cost of advertising may preclude large-scale campaigns. In addition, viral campaigns to increase awareness of internet interventions on social media may be considered. When delivering internet interventions in a community setting, inclusion of these features may lead to greater uptake and adherence.

Limitations

There were many limitations to this study. The sample was limited to a single state in Australia, with predominantly highly educated (university students) participants. In addition, overall, very few males participated, with none responding to the EFI evaluation survey. However, this is common in other self-selected samples for studies on anxiety and depression [52] where females are overrepresented. In addition, the sample reflects similar gender imbalances in the usage of e–mental health programs in the community, where substantially more females are found to use these services [24]. Engaging men in mental health research is challenging; they may be less likely than women to participate because of specific factors such as stigma [53]. Despite this, these issues with the sample limit the generalizability of the study to other populations. It is also possible that the evaluation survey could have been biased toward those who were more engaged in the process, and thus, this may not be reflective of the entire sample. In addition, many participants stated that they believed the resultant EFI would increase their engagement with a web-based self-help e–mental health program; however, this may not reflect their actual behavior.

Conclusions

One of the critical barriers to the uptake of e–mental health programs is the lack of certainty among community members about whether the program can help them. A second major barrier to uptake is the lack of awareness of e–mental health

programs and their availability in the community. This study found that an EFI for depression and anxiety primarily requires information about what the program involves, such as evidence for its effectiveness, normalization of participation in e–mental health programs, including testimonials, and finally, brief information on data security, although this factor was not as prevalent as expected. Attention to these factors may guide the development of future technology-based interventions that are designed to increase engagement and adherence.

Acknowledgments

The authors wish to thank Nazrana Khaled for assistance with notetaking for two of the focus groups. This project was supported by a National Health and Medical Research Council grant (number 1138713). PB is supported by National Health and Medical Research Council (NHMRC) Fellowship 1158707; AC is supported by NHMRC Fellowship 1122544; FK is supported by NHMRC Fellowship 1110371; LF is supported by Australian Research Council Discovery Early Career Researcher Award 190101382; and MB is supported by Medical Research Future Fund Fellowship 1150698.

Authors' Contributions

AG drafted the manuscript, conducted the study, and performed the analysis. PB conceived the study and supervised the study. PB, AC, FK, and MS provided input into the study design and secured the funding. LF provided clinical support and assisted with conducting the groups. MB assisted with methodology and provided advice on qualitative analysis. All authors reviewed and critically edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reporting checklist for qualitative study.
[DOCX File , 25 KB - formative_v4i10e22528_app1.docx]

Multimedia Appendix 2
List of questions.
[DOCX File , 18 KB - formative_v4i10e22528_app2.docx]

Multimedia Appendix 3
Written activities.
[DOCX File , 18 KB - formative_v4i10e22528_app3.docx]

Multimedia Appendix 4
Questions from online follow-up survey.
[DOCX File , 14 KB - formative_v4i10e22528_app4.docx]

Multimedia Appendix 5
Draft EFI presented to participants in evaluation survey.
[PDF File (Adobe PDF File), 971 KB - formative_v4i10e22528_app5.pdf]

References

1. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013;34:119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
2. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci Rep 2018 Feb 12;8(1):2861 [FREE Full text] [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
3. Baxter AJ, Scott KM, Vos T, Whiteford HA. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol Med 2013 May;43(5):897-910. [doi: 10.1017/S003329171200147X] [Medline: 22781489]
4. Lee Y, Rosenblat JD, Lee J, Carmona NE, Subramaniapillai M, Shekotikhina M, et al. Efficacy of antidepressants on measures of workplace functioning in major depressive disorder: A systematic review. J Affect Disord 2018 Feb;227:406-415. [doi: 10.1016/j.jad.2017.11.003] [Medline: 29154157]

5. Hendriks SM, Spijker J, Licht CM, Hardeveld F, de Graaf R, Batelaan NM, et al. Long-term work disability and absenteeism in anxiety and depressive disorders. J Affect Disord 2015 Jun 1;178:121-130. [doi: 10.1016/j.jad.2015.03.004] [Medline: 25805404]
6. Burgess PM, Pirkis JE, Slade TN, Johnston AK, Meadows GN, Gunn JM. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2009 Jul;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
7. Zhang M, Cheow E, Ho CS, Ng BY, Ho R, Cheok CC. Application of low-cost methodologies for mobile phone app development. JMIR Mhealth Uhealth 2014 Dec 9;2(4):e55 [FREE Full text] [doi: 10.2196/mhealth.3549] [Medline: 25491323]
8. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
9. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Sep;49(9):776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
10. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
11. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
12. Ebert D, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 1;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
13. Apolinário-Hagen J, Harrer M, Kählke F, Fritsche L, Salewski C, Ebert DD. Public attitudes toward guided internet-based therapies: web-based survey study. JMIR Ment Health 2018 May 15;5(2):e10735 [FREE Full text] [doi: 10.2196/10735] [Medline: 29764797]
14. Mohr DC, Siddique J, Ho J, Duffecy J, Jin L, Fokuo JK. Interest in behavioral and psychological treatments delivered face-to-face, by telephone, and by internet. Ann Behav Med 2010 Aug;40(1):89-98 [FREE Full text] [doi: 10.1007/s12160-010-9203-7] [Medline: 20652466]
15. Casey LM, Wright M, Clough BA. Comparison of perceived barriers and treatment preferences associated with internet-based and face-to-face psychological treatment of depression. Int J Cyber Behav Psychol 2014 Oct;4(4):16-22. [doi: 10.4018/ijcbpl.2014100102]
16. Clough BA, Zarean M, Ruane I, Mateo NJ, Aliyeva TA, Casey LM. Going global: do consumer preferences, attitudes, and barriers to using e-mental health services differ across countries? J Ment Health 2019 Feb;28(1):17-25. [doi: 10.1080/09638237.2017.1370639] [Medline: 28857650]
17. Choi I, Sharpe L, Li S, Hunt C. Acceptability of psychological treatment to Chinese- and Caucasian-Australians: internet treatment reduces barriers but face-to-face care is preferred. Soc Psychiatry Psychiatr Epidemiol 2015 Jan;50(1):77-87. [doi: 10.1007/s00127-014-0921-1] [Medline: 24993290]
18. Head to Health. URL: https://headtohealth.gov.au/ [accessed 2018-10-11] [WebCite Cache ID 7354K9g5n]
19. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
20. Australian Government Response to Contributing Lives, Thriving Communities: Review of Mental Health Programmes and Services. Australian Government. 2015. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/mental-review-response [accessed 2020-10-16]
21. Gun SY, Titov N, Andrews G. Acceptability of internet treatment of anxiety and depression. Australas Psychiatry 2011 Jun;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]
22. Young KS. An empirical examination of client attitudes towards online counseling. Cyberpsychol Behav 2005 Apr;8(2):172-177. [doi: 10.1089/cpb.2005.8.172] [Medline: 15938657]
23. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
24. Batterham P, Neil A, Bennett K, Griffiths K, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 2;2:97-105 [FREE Full text] [Medline: 19920949]
25. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

26. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
27. Piotrowska PJ, Tully LA, Lenroot R, Kimonis E, Hawes D, Moul C, et al. Mothers, fathers, and parental systems: a conceptual model of parental engagement in programmes for child mental health-connect, attend, participate, enact (CAPE). Clin Child Fam Psychol Rev 2017 Jun;20(2):146-161 [FREE Full text] [doi: 10.1007/s10567-016-0219-9] [Medline: 27914017]
28. Ingoldsby EM. Review of interventions to improve family engagement and retention in parent and child mental health programs. J Child Fam Stud 2010 Oct 1;19(5):629-645 [FREE Full text] [doi: 10.1007/s10826-009-9350-2] [Medline: 20823946]
29. Ajzen I. The theory of planned behavior. Organ Behav Hum Dec 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
30. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
31. Baumeister H, Seifferth H, Lin J, Nowoczin L, Lüking M, Ebert D. Impact of an acceptance facilitating intervention on patients' acceptance of internet-based pain interventions: a randomized controlled trial. Clin J Pain 2015 Jun;31(6):528-535. [doi: 10.1097/AJP.0000000000000118] [Medline: 24866854]
32. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2018 Aug 21;20(8):e244 [FREE Full text] [doi: 10.2196/jmir.9925] [Medline: 30131313]
33. Gerhards S, Abma T, Arntz A, de Graaf L, Evers S, Huibers M, et al. Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. J Affect Disord 2011 Mar;129(1-3):117-125. [doi: 10.1016/j.jad.2010.09.012] [Medline: 20889214]
34. Miller CL, Mott K, Cousins M, Miller S, Johnson A, Lawson T, et al. Integrating consumer engagement in health and medical research - an Australian framework. Health Res Policy Syst 2017 Feb 10;15(1):9 [FREE Full text] [doi: 10.1186/s12961-017-0171-2] [Medline: 28187772]
35. Suomi A, Freeman B, Banfield M. Framework for the Engagement of People With a Lived Experience in Program Implementation and Research. Black Dog Institute. 2017. URL: https://www.blackdoginstitute.org.au/wp-content/uploads/2020/04/anu-lived-experience-framework.pdf [accessed 2020-10-16]
36. Bovaird T. Beyond engagement and participation: user and community coproduction of public services. Public Admin Rec 2007 Sep;67(5):846-860. [doi: 10.1111/j.1540-6210.2007.00773.x]
37. Batterham PJ, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Gulliver A. A brief intervention to increase uptake and adherence of an online program for depression and anxiety: Protocol for the Enhancing Engagement with Psychosocial Interventions (EEPI) Randomized Controlled Trial. Contemp Clin Trials 2019 Mar;78:107-115. [doi: 10.1016/j.cct.2019.01.015] [Medline: 30711664]
38. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
39. Kitzinger J. The methodology of Focus Groups: the importance of interaction between research participants. Sociol Health Illness 1994 Jan;16(1):103-121. [doi: 10.1111/1467-9566.ep11347023]
40. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016 Jun;38(3):655-662 [FREE Full text] [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]
41. Gill PJ, Hewitson P, Peile E, Harnden A. Prioritizing areas for quality marker development in children in UK general practice: extending the use of the nominal group technique. Fam Pract 2012 Oct;29(5):567-575. [doi: 10.1093/fampra/cms006] [Medline: 22308179]
42. van Rijnsoever FJ. (I Can't Get No) Saturation: a simulation and guidelines for sample sizes in qualitative research. PLoS One 2017;12(7):e0181689 [FREE Full text] [doi: 10.1371/journal.pone.0181689] [Medline: 28746358]
43. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Appiah O. Rich media, poor media: the impact of audio/video vs text/picture testimonial ads on browsers'evaluations of commercial web sites and online products. J Curr Issues Res Advert 2006 Mar;28(1):73-86. [doi: 10.1080/10641734.2006.10505192]
45. Greene K, Brinn LS. Messages influencing college women's tanning bed use: statistical versus narrative evidence format and a self-assessment to increase perceived susceptibility. J Health Commun 2003;8(5):443-461. [doi: 10.1080/713852118] [Medline: 14530147]
46. Neighbors C, Lewis MA, Atkins DC, Jensen MM, Walter T, Fossos N, et al. Efficacy of web-based personalized normative feedback: a two-year randomized controlled trial. J Consult Clin Psychol 2010 Dec;78(6):898-911 [FREE Full text] [doi: 10.1037/a0020766] [Medline: 20873892]
47. Hom M, Heaney C, Koopman C. Personalized normative feedback for depression symptoms: a qualitative pilot study of female undergraduates. Acad Psychiatry 2014 Aug;38(4):464-469. [doi: 10.1007/s40596-014-0076-0] [Medline: 24664603]

48. Lewis MA, Neighbors C. Social norms approaches using descriptive drinking norms education: a review of the research on personalized normative feedback. J Am Coll Health 2006;54(4):213-218 [FREE Full text] [doi: 10.3200/JACH.54.4.213-218] [Medline: 16450845]
49. Reynolds J, Griffiths KM, Cunningham JA, Bennett K, Bennett A. Clinical practice models for the use of e-mental health resources in primary health care by health professionals and peer workers: a conceptual framework. JMIR Ment Health 2015;2(1):e6 [FREE Full text] [doi: 10.2196/mental.4200] [Medline: 26543912]
50. Batterham PJ, Calear AL, Gulliver A, Farrer LM. Efficacy of a transdiagnostic, video-based online program for reducing depression, anxiety, and suicidal ideation in adults: Protocol for a randomised controlled trial. Contemp Clin Trials Commun 2019 Jun;14:100341 [FREE Full text] [doi: 10.1016/j.conctc.2019.100341] [Medline: 30911696]
51. Banfield MA, Morse AR, Gulliver A, Griffiths KM. Mental health research priorities in Australia: a consumer and carer agenda. Health Res Policy Syst 2018 Dec 12;16(1):119 [FREE Full text] [doi: 10.1186/s12961-018-0395-9] [Medline: 30541546]
52. Batterham PJ, Sunderland M, Carragher N, Calear AL, Mackinnon AJ, Slade T. The distress questionnaire-5: population screener for psychological distress was more accurate than the K6/K10. J Clin Epidemiol 2016 Mar;71:35-42. [doi: 10.1016/j.jclinepi.2015.10.005] [Medline: 26464194]
53. Woodall A, Morgan C, Sloan C, Howard L. Barriers to participation in mental health research: are there specific gender, ethnicity and age related barriers? BMC Psychiatry 2010 Dec 2;10:103 [FREE Full text] [doi: 10.1186/1471-244X-10-103] [Medline: 21126334]

Abbreviations

AFI: acceptance-facilitation intervention
EFI: engagement-facilitation intervention
NHMRC: National Health and Medical Research Council

Edited by G Eysenbach; submitted 15.07.20; peer-reviewed by A AL-Asadi, T Nygren; comments to author 25.08.20; revised version received 28.08.20; accepted 02.10.20; published 29.10.20.

Please cite as:
Gulliver A, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Banfield M, Batterham PJ
Consumer-Guided Development of an Engagement-Facilitation Intervention for Increasing Uptake and Adherence for Self-Guided Web-Based Mental Health Programs: Focus Groups and Online Evaluation Survey
JMIR Form Res 2020;4(10):e22528
URL: http://formative.jmir.org/2020/10/e22528/
doi:10.2196/22528
PMID:33118939

Original Paper

Identification of Emotional Expression With Cancer Survivors: Validation of Linguistic Inquiry and Word Count

Michelle McDonnell[1], PhD; Jason Edward Owen[2], PhD; Erin O'Carroll Bantum[3], PhD

[1]Veteran's Affairs Loma Linda Healthcare System, Loma Linda, CA, United States

[2]US Department of Veterans Affairs, National Center for PTSD, VA Palo Alto Health Care System, Palo Alto, CA, United States

[3]Cancer Prevention in the Pacific, University of Hawaii Cancer Center, Honolulu, HI, United States

Corresponding Author:
Erin O'Carroll Bantum, PhD
Cancer Prevention in the Pacific
University of Hawaii Cancer Center
701 Ilalo St, B4
Honolulu, HI, 96813
United States
Phone: 1 8084413491
Email: ebantum@cc.hawaii.edu

Abstract

Background: Given the high volume of text-based communication such as email, Facebook, Twitter, and additional web-based and mobile apps, there are unique opportunities to use text to better understand underlying psychological constructs such as emotion. Emotion recognition in text is critical to commercial enterprises (eg, understanding the valence of customer reviews) and to current and emerging clinical applications (eg, as markers of clinical progress and risk of suicide), and the Linguistic Inquiry and Word Count (LIWC) is a commonly used program.

Objective: Given the wide use of this program, the purpose of this study is to update previous validation results with two newer versions of LIWC.

Methods: Tests of proportions were conducted using the total number of emotion words identified by human coders for each emotional category as the reference group. In addition to tests of proportions, we calculated F scores to evaluate the accuracy of LIWC 2001, LIWC 2007, and LIWC 2015.

Results: Results indicate that LIWC 2001, LIWC 2007, and LIWC 2015 each demonstrate good sensitivity for identifying emotional expression, whereas LIWC 2007 and LIWC 2015 were significantly more sensitive than LIWC 2001 for identifying emotional expression and positive emotion; however, more recent versions of LIWC were also significantly more likely to overidentify emotional content than LIWC 2001. LIWC 2001 demonstrated significantly better precision (F score) for identifying overall emotion, negative emotion, and anxiety compared with LIWC 2007 and LIWC 2015.

Conclusions: Taken together, these results suggest that LIWC 2001 most accurately reflects the emotional identification of human coders.

(JMIR Form Res 2020;4(10):e18246) doi:10.2196/18246

KEYWORDS

linguistic analysis; emotion; validation

Introduction

Recent studies have provided evidence that emotions can be effectively identified in written text [1-4]. Written emotions have been identified as significantly different from characteristically nonemotional writing, such as academic tasks [4], and more importantly, they can be correctly identified by readers [3]. As the use of web-based interventions, mobile app, social media, other text communication (eg, emails and text messages), and web-based communication (eg, Zoom and Skype) increases, the need for validated tests that can rapidly analyze text data has become exceptionally important. A large proportion of emotion-based text data are currently analyzed by human coders [5,6] as they add predictive power above and beyond that of computer analyses for the identification of positive emotions [7] as well as symptoms of depression [8].

Although computer analysis has become increasingly more efficient in evaluating written text, it lacks the nuance and accuracy provided by human coders [9]. Although qualitative analysis provides the most complete method for characterizing text-based communications [10], the cost, time requirements, and subjectivity of manual coding make these methods prohibitively difficult for many applications. Computerized text analysis programs may be able to ameliorate these limitations. Unfortunately, many computerized text analysis programs are not validated for the identification of emotional expression and still have time-consuming data preparation to ensure that the text is clear of all typographical errors [9].

The Linguistic Inquiry and Word Count (LIWC) [11] is a text analysis program that calculates the degree to which various categories of words are used in text. Bantum and Owen [12] previously evaluated the validity of the LIWC 2001, demonstrating that LIWC 2001 had good sensitivity and specificity for identifying emotion; however, the positive predictive value (PPV), or precision of emotional identification, was poor. Additional work by our team with the creation of a machine learning program [13] demonstrated that a machine learning approach was not necessarily more predictive than LIWC. Since our previous validation study, Pennebaker et al [14] have released 2 updates: LIWC 2007 and LIWC 2015, both of which have removed categories found to have poor emotional identification (eg, optimism) and increased the dictionary size of the program. These updates have greatly altered the dictionary for this program, including new categories of drive, analytical thinking, emotional tones, time orientation, and relativity and new subcategories of gender references and netspeak. In terms of emotional content, the largest change involves the inclusion of *netspeak* to quantify emotional expression (eg, ":)" would be categorized as a positive emotion word).

It is essential to validate LIWC 2007 and LIWC 2015 due to the widespread use of LIWC in research, clinical treatment, and commercial applications. LIWC 2001 and 2015 successfully identified individuals at risk for suicide as they had an increased presence of first-person singular self-references and negative emotional expression as they approached their suicide [1,2,15]. Sonnenschein et al [16] indicated LIWC 2007 noted that depressed individuals produced more words related to the sadness subcategory when compared with anxious individuals. Overall, LIWC has been determined to be able to accurately identify language differences between people who are diagnosed with attention deficit hyperactivity disorder, anxiety, bipolar, borderline, depression, eating disorders, obsessive-compulsive disorder, schizophrenia, and seasonal affective disorder [17,18]. LIWC 2007 has been used to evaluate whether a clinician's distancing language increases suicide rates among veterans [19], assess what men and women value in romantic relationships [20], determine how social media predicts the outcomes of presidential elections [21], and evaluate language style and its ability to determine relationship initiation and stability [22].

Similar to LIWC 2001, the original evaluation of the validity of LIWC 2007 and LIWC 2015 was based on the results of a series of correlations between judges' ratings and LIWC scores. Bantum and Owen [12] found that LIWC have low predictive value and overidentification of emotions, despite significant, though modest, correlations between judges' ratings and LIWC scores [12]. Given these shortcomings and the need to independently validate LIWC 2007 and LIWC 2015 for emotion recognition in text [23,24], we sought to determine whether LIWC 2007 and LIWC 2015 are valid for emotion recognition in text and whether they improve upon the known limitations of LIWC 2001. The first aim of this study was to assess the accuracy of LIWC 2007 and LIWC 2015 for the detection of emotional expression using tests of specificity, sensitivity, PPV, and negative predictive value (NPV). The second aim was to evaluate potential differences between LIWC 2001, LIWC 2007, and LIWC 2015 for emotional identification. It is hypothesized that LIWC 2015 will be more efficacious in emotional identification than LIWC 2007 and LIWC 2001 because of the significant alterations in the most recent dictionary, recent regard for some contextual information in written text, and continued improvement in word selection through the construction of the product.

Methods

Participants

The participants were recruited from a hematology/oncology outpatient clinic at a large medical center in the southeastern United States. The participants included 49 women with stage 1 or stage 2 breast cancer and 14 women with stage 3 or stage 4 breast cancer. Participants were not excluded from participation based on the time elapsed since their diagnosis or medical treatment. The women participated in a randomized 12-week clinical trial of an internet-based support group. Additional information regarding the sample has been previously reported [25]. The textual data of the 63 participants were analyzed for this particular study. The participants had a mean age of 49.8 years (SD 11.0), the majority were college educated (mean 15.4 years, SD 2.4), and they were largely of White descent (57/63, 93%).

Procedures

All participants completed a baseline assessment once they agreed to participate in the study, before being given access to the web-based support group. Once the participants were given access to the web-based support group, they were encouraged to communicate with one another through a discussion board regarding general topics and a series of interactive coping-skills training. The textual data were stored in an individual data file for each participant. Further information regarding the experimental procedures for these participants has been previously reported [12,26]. The study data set is identical to that used in the original validation study of LIWC 2001.

Rater Coding of Emotional Expression

This particular study utilized human-coded ratings of emotion generated in a previous analysis of these linguistic data [12]. To briefly describe how these codes were generated, Bantum and Owen [12] had a well-defined set of rater coding rules for the human coders to follow. Each coder independently identified all words containing emotional expression. If the coders determined that emotional expression was present in a given sentence (as taken into consideration within the context of the

text), then the word that most acutely captured the emotion was placed into the best fitting subcategory, based on categories specified by LIWC 2001. Eight potential categories were used: "positive feelings," "optimism," "anxiety," "anger," "sadness," "other positive emotion," "other negative emotion," or "not emotion." Coders were asked to differentiate between what was likely to be a physical pain and emotional pain or experience (eg, "the impact of chemotherapy on me physically was physically exhausting"). This process took quite a lot of training and judgment, as the use of emotion can be ambiguous, especially solely in text. Any discrepancies found between coders were discussed among the researchers, and final codes were established by consensus. The interrater reliability between the two trained coders was excellent (κ=0.80). Two additional raters were trained on the coding process and then reviewed 33% of the text. The interrater reliability was evaluated between the two additional raters and was established to have substantial agreement between raters (κ=0.69).

Materials

LIWC 2001

LIWC 2001 is a computational text analysis method that processes text-based data on a word-by-word basis. LIWC compares each word of text with a library of 5 categories (ie, linguistic dimensions, psychological processes, relativity, personal concerns, and experimental dimensions) of words to identify whether each specific word from the source data set matches any of the words or word fragments found in the LIWC library. With regard to emotion, words are compared with each of the 3 categories of emotion (eg, emotional expression, positive emotion, and negative emotion). For each word that is established as a match to a word or fragment in a LIWC emotion dictionary, the program iterates a count of all emotion words identified in that particular emotion dictionary (eg, positive emotion). LIWC uses the results of the word count to establish a percentage of total words in the text that contain emotion words or a specific type of emotion. After the word has been identified as positive or negative, it is then placed into a specific category, such as positive feelings or optimism in the positive category and sadness, anger, or anxiety in the negative category. In some instances, words may be identified as expressing emotion and categorized as positive or negative, but LIWC is unable to resolve a more specific emotion category (eg, anxiety vs sadness), which may cause it to identify a word as belonging to multiple emotion dictionaries.

LIWC 2007

Each participant's text information was also analyzed using LIWC 2007 (n=63) [14]. LIWC 2007 has a similar structure to that of LIWC 2001 in that it is a computational text analysis program that evaluates each item on a word-by-word basis. Furthermore, LIWC 2007 also provides a percentage of total words that are represented by emotion.

Pennebaker et al [14] developed LIWC 2007 to specifically address a number of key limitations in LIWC 2001, such as a limited dictionary, uncommonly used word categories, and a lack of function words (eg, conjunctions, adverbs, quantifiers, auxiliary verbs, and impersonal pronouns). The creators of LIWC 2007 removed the following word categories found in LIWC 2001 because they had poor base rates: optimism, positive feelings, communication verbs, metaphysical, sleeping, grooming, and school. The new dictionary for LIWC 2007 was altered to provide more accurate word categories by omitting those categories with insufficient validity and adding a number of categories to represent function words as well as including previously experimental categories into the program (eg, swear words, nonfluencies, and fillers). Additionally, researchers increased the dictionary count from 2300 words and word stems to 4500 words so that it may better represent emotional expression and other key psychological constructs. There were a number of words defined as emotion in the LIWC 2007 dictionary that were previously categorized as nonemotion (eg, confident, champ, and resolve). In addition to a reclassification of preexisting words, LIWC 2007 added emotional words that were not originally included in the LIWC 2001 dictionary (eg, grace, jaded, joke, openness, and rancid) and removed some emotional words that were in the LIWC 2001 dictionary (eg, sensitive). Finally, the LIWC 2007 dictionary classified the roots of words as emotion (eg, stammer) that may be perceived as nonemotion by human coders in an extended form (eg, stammered and stammering).

LIWC 2015

Each participant's text information was also analyzed using LIWC 2015 (n=63) [27]. LIWC 2015 has a similar structure to the previous versions in that it evaluates textual data on a word-by-word basis, looking for specific target words within the appropriate word category scales. As words are identified as target words, they are classified into one or more categories or subcategories, which are then arranged hierarchically. Specifically, the word "cried" may be placed in the categories of sadness, negative emotion, overall affect, verbs, and past focus. Furthermore, LIWC 2015 provides a percentage of total words that are represented by emotion, and various structural composition elements (eg, sentence punctuation) are also incremented.

Pennebaker et al [27] created an entirely new dictionary and software system, rather than simply updating the previous versions of LIWC. This resulted in an increase in the dictionary size, totaling nearly 6400 words, word stems, and select emoticons. In addition to an increase in dictionary size, LIWC 2015 added 9 word categories tapping into psychological constructs (eg, affect and cognition) and 5 informal language categories (assents, fillers, swear words, and netspeak). There was the removal of one personal concern category (eg, work, home, and leisure activities) and one standard linguistic category (eg, percentage of words in the text that are pronouns, articles, and adjectives). Pennebaker et al [27] also reclassified how the word "like" is categorized in an attempt to reduce the risk of false positives that may occur in utterances, comparatives, or prepositions. Specifically, in previous versions of LIWC, the word "like" was categorized as indicative of emotion, particularly positive emotion. In LIWC 2015, the word "like" only qualifies as an emotion word when attributed to a person or an action, such as "I like," "they like," and "will like." Pennebaker et al [27] added a category of *Netspeak* to assess for words and communication styles (eg, btw, lol, and thx)

common on social media platforms (eg, Facebook, Snapchat, and instant messaging).

Data Preparation

Each time the individuals participated in the web-based support group, their textual data were saved in person-specific files. The files were then combined into a single spreadsheet per participant so that each word was considered a subject. The text files from the human rater coding of emotional expression were created and combined with the results from the LIWC 2001, LIWC 2007, and LIWC 2015 data analysis. Each instance of emotion was counted as one point, and the frequency of a given emotion was divided by total words for that participant, resulting in a percentage of a given emotion for each participant. This was true for LIWC 2001, LIWC 2007, and LIWC 2015.

Data Analytic Plan

This study contains a total of 165,754 words consisting of 278 single-spaced, 12-point font pages. Each word is considered as a single variable. An analysis of power was conducted using G*Power 3 [28] and indicated that with an effect size of 0.5, alpha level of .01, and a sample size of 165,754, the power to detect between-method differences in greater than 0.80. To assess the differences between LIWC 2001, LIWC 2007, and LIWC 2015 with regard to the accuracy of emotional identification, we calculated tests of proportions. We calculated the sensitivity, specificity, PPV, and NPV to identify the proportion of words that were similarly identified by human coders for LIWC 2001, LIWC 2007, and LIWC 2015. Subsequently, we utilized the overall proportions for each emotional category and conducted the test of proportions using the total number of emotion words identified by human coders for each emotional category as the reference group. To control for the number of tests, we calculated a Bonferroni correction for the *P* value to provide more stringent criteria for meeting sensitivity, which was calculated at *P*=.008. In addition to tests of proportions, we calculated F scores to evaluate the accuracy of LIWC 2001, LIWC 2007, and LIWC 2015. Accuracy was assessed by considering the harmonic mean of precision, or the fraction of retrieved items that are relevant, and recall, or the fraction of retrieved items that are relevant and successfully retrieved, of each program. This is a way of balancing the various measures (sensitivity and PPV) that we have evaluated. The F score was calculated by multiplying 2 with the results of PPV (precision) multiplied by sensitivity (recall) divided by PPV (precision) added to sensitivity (recall).

Results

Overview

The average percentage of words identified by LIWC 2001, LIWC 2007, LIWC 2015, and human coders as emotion, positive emotion, and negative emotion as well as specific subcategories of anxiety, anger, and sadness ranged from 0.1% to 4.1% of total words (Table 1). Pearson correlations were conducted to evaluate the strength of the relationship between each version of LIWC as well as human coders (Tables 2 and 3). The results revealed significant positive relationships across each version of LIWC, particularly between LIWC 2007 and LIWC 2015 (affect, r=0.95, P<.01; positive, r=0.95, P<.01; negative, r=0.97, P<.01; anxiety, r=0.97, P<.01; anger, r=0.91, P<.01; and sad, r=0.90, P<.01). With regard to the relationship between LIWC and human coders, the largest positive relationships occurred between LIWC 2001 (positive, r=0.47, P<.01; negative emotions, r=0.58, P<.01; and anxiety, r=0.74, P<.01) and LIWC 2015 (affect, r=0.55, P<.01; anger, r=0.51, P<.01; and sad, r=0.57, P<.01).

Table 1. Linguistic Inquiry and Word Count 2001, 2007, 2015, and human coder average of words identified in each emotional category and subcategory.

Type of emotion	LIWC[a] 2001, % (SD)	LIWC 2007, % (SD)	LIWC 2015, % (SD)	Human coders, % (SD)
Total affect	4.8 (0.214)	6.1 (0.239)	6.1 (0.239)	1.8 (0.134)
Positive emotion	3.2 (0.175)	4.1 (0.198)	4.1 (0.198)	0.9 (0.096)
Negative emotion	1.6 (0.125)	1.9 (0.137)	1.9 (0.138)	0.9 (0.094)
Anxiety	0.5 (0.067)	0.6 (0.079)	0.6 (0.077)	0.3 (0.058)
Anger	0.2 (0.046)	0.2 (0.050)	0.2 (0.046)	0.1 (0.034)
Sadness	0.3 (0.059)	0.4 (0.063)	0.2 (0.066)	0.2 (0.044)

[a]LIWC: Linguistic Inquiry and Word Count.

Table 2. Correlations between Linguistic Inquiry and Word Count 2001, 2007, and 2015 with human codes for classification of emotion.

Type of emotion	Human codes					
	Affect	Positive	Negative	Anxiety	Anger	Sad
LIWC[a] 2001						
Affect	0.50[b]	0.38[b]	0.34[b]	0.23[b]	0.12[b]	0.17[b]
Positive	−0.03[b]	0.47[b]	0.00	−0.01[b]	0.00	−0.01
Negative	0.89[b]	0.00	0.58[b]	0.40[b]	0.20[b]	0.30[b]
Anxiety	0.47[b]	0.01[b]	0.54[b]	0.74[b]	0.02[b]	−0.00
Anger	0.33[b]	−0.00	0.23[b]	0.04[b]	0.49[b]	0.00
Sad	0.42[b]	0.01	0.29[b]	0.00	0.00	0.52[b]
LIWC 2007						
Affect	0.53[b]	0.35[b]	0.32[b]	0.21[b]	0.11[b]	0.16[b]
Positive	−0.03[b]	0.43[b]	−0.00	−0.01[b]	−0.00	−0.01[b]
Negative	1.00[c]	0.00	0.55[b]	0.38[b]	0.19[b]	0.28[b]
Anxiety	0.56[b]	0.01[b]	0.50[b]	0.65[b]	0.02[b]	−0.00
Anger	0.34[b]	−0.01	0.19	0.00	0.46[b]	0.00
Sad	0.45[b]	0.01[c]	0.28[b]	0.00	0.00	0.50[b]
LIWC 2015						
Affect	0.55[b]	0.36[b]	0.35[b]	0.21[b]	0.11[b]	0.15[b]
Positive	−0.30	0.43[b]	−0.00	−0.01[b]	−0.00	0.01[b]
Negative	1.00[b]	0.00	0.54[b]	0.37[b]	0.20[b]	0.27[b]
Anxiety	0.55[b]	0.01[b]	0.49[b]	0.66[b]	0.02[b]	−0.00
Anger	0.33[b]	−0.00	0.21[b]	−0.00	0.51[b]	0.00
Sad	0.47[b]	0.01[c]	0.31[b]	0.00	0.00	0.57[b]

[a]LIWC: Linguistic Inquiry and Word Count.

[b]$P<.001$.

Table 3. Correlation matrix comparing Linguistic Inquiry and Word Count 2001, 2007, and 2015.

Type of emotion	2001						2007						2015					
	Affect	Positive	Negative	Anxiety	Anger	Sad	Affect	Positive	Negative	Anxiety	Anger	Sad	Affect	Positive	Negative	Anxiety	Anger	Sad
2001																		
Affect	—[a]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Positive	0.81[b]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Negative	0.57[b]	−0.02[b]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Anxiety	0.30[b]	−0.01[b]	0.53[b]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Anger	0.21[b]	−0.01[b]	0.36[b]	0.05[b]	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Sad	0.26[b]	−0.01[b]	0.46[b]	0.03[b]	−0.00	—	—	—	—	—	—	—	—	—	—	—	—	—
2007																		
Affect	0.86[b]	0.70[b]	0.50[b]	0.27[b]	0.18[b]	0.23[b]	—	—	—	—	—	—	—	—	—	—	—	—
Positive	0.69[b]	0.86[b]	−0.03[b]	−0.01[b]	−0.01[b]	−0.01[b]	0.81[b]	—	—	—	—	—	—	—	—	—	—	—
Negative	0.51[b]	−0.03[b]	0.91[b]	0.48[b]	0.33[b]	0.42[b]	0.55[b]	−0.03[b]	—	—	—	—	—	—	—	—	—	—
Anxiety	0.31[b]	−0.01[b]	0.56[b]	0.85[b]	0.04[b]	0.02[b]	0.31[b]	−0.02[b]	0.57[b]	—	—	—	—	—	—	—	—	—
Anger	0.19[b]	−0.01[b]	0.33[b]	−0.00	.086[b]	−0.00	0.20[b]	−0.01[b]	0.36[b]	−0.00	—	—	—	—	—	—	—	—
Sad	0.26[b]	−0.01[b]	0.46[b]	0.02[b]	−0.00	0.93[b]	0.25[b]	−0.01[b]	0.45[b]	0.02[b]	−0.03	—	—	—	—	—	—	—
2015																		
Affect	0.85[b]	0.68[b]	0.49[b]	0.26[b]	0.18[b]	0.23[b]	0.95[b]	0.77[b]	0.54[b]	0.31[b]	0.19[b]	0.25[b]	—	—	—	—	—	—
Positive	0.67[b]	0.84[b]	−0.03[b]	−0.01[b]	−0.01[b]	−0.01[b]	0.77[b]	0.95[b]	−0.03[b]	−.02[b]	−0.01[b]	−0.01[b]	0.81[b]	—	—	—	—	—
Negative	0.50[b]	−0.03[b]	0.89[b]	0.47[b]	0.33[b]	0.42[b]	0.53[b]	−0.03[b]	0.97[b]	0.56[b]	0.34[b]	0.45[b]	0.55[b]	−0.03[b]	—	—	—	—
Anxiety	0.31[b]	−0.01[b]	0.54[b]	0.82[b]	0.05[b]	0.02[b]	0.31[b]	−0.02[b]	0.56[b]	0.97[b]	0.01[b]	0.02[b]	0.31[b]	−0.02[b]	0.55[b]	—	—	—
Anger	0.18[b]	−.01[b]	0.32[b]	−0.00	0.80[b]	−0.00	0.18[b]	−0.01[b]	0.32[b]	−0.00	0.91[b]	−0.00	0.18[b]	−0.01[b]	0.33[b]	0.00	—	—
Sad	0.28[b]	−0.01[b]	0.50[b]	0.02[b]	−0.00	0.88[b]	0.26[b]	−0.01[b]	0.47[b]	0.02	−0.00	0.90[b]	0.26[b]	−0.01[b]	0.47[b]	0.02[b]	−0.00	—

[a]Not applicable.

[b]$P<.01$.

Relationship Between LIWC 2001 and LIWC 2007 Coding Methods

Sensitivity

Sensitivity captured the proportion of total emotion words identified by human raters as being representative of emotional expression that were also captured by LIWC 2001, LIWC 2007, or LIWC 2015. Sensitivity for overall emotional expression was good (eg, greater than 0.80, equivalent to at least 80% accuracy) for all 3 versions: LIWC 2001 (0.858), LIWC 2007 (0.896), and LIWC 2015 (0.904; Table 4). Additionally, all 3 versions demonstrated good sensitivity for positive emotion, negative emotion, and anxiety, with results ranging from 0.819 to 0.892. Alternatively, LIWC 2001 and LIWC 2007 demonstrated poor (less than 0.80) performance in anger (0.663 and 0.679, respectively) and sadness (0.699 and 0.718, respectively), whereas LIWC 2015 demonstrated good performance for sadness (0.856) and poor performance for anger (0.695). Compared with LIWC 2001, LIWC 2007 and LIWC 2015 produced significantly higher identification of emotional

expression in the categories of overall emotion and positive emotion. Additionally, LIWC 2015 demonstrated higher sensitivity for the category of sadness when compared with LIWC 2001 and LIWC 2007. Notably, there were no differences between LIWC 2001, LIWC 2007, and LIWC 2015 for negative emotions, anxiety, or anger.

Table 4. Linguistic Inquiry and Word Count 2001, 2007, and 2015 sensitivity and specificity with 95% CI (N=63)[a].

Emotion categories	2001 sensitivity (95% CI)	2007 sensitivity (95% CI)	2015 sensitivity (95% CI)	2001 specificity (95% CI)	2007 specificity (95% CI)	2015 specificity (95% CI)
Total affect	0.858[a,b] (0.845-0.871)	0.896[b] (0.884-0.906)	0.904[c] (0.894-0.915)	0.967 (0.966-0.968)	0.955 (0.954-0.956)	0.955 (0.954-0.956)
Positive emotion	0.873[b,c] (0.855-0.889)	0.913[b] (0.898-0.927)	0.928 (0.915-0.941)	0.976 (0.975-0.977)	0.967 (0.966-0.967)	0.967 (0.958-0.976)
Negative emotion	0.822 (0.803-0.839)	0.814 (0.793-0.834)	0.810 (0.789-0.830)	0.990 (0.990-0.991)	0.987 (0.987-0.988)	0.988 (0.982-0.994)
Anxiety	0.862 (0.829-0.888)	0.892 (0.863-0.916)	0.883 (0.856-0.910)	0.998 (0.998-0.999)	0.997 (0.996-0.997)	0.997 (0.993-1.00)
Anger	0.663 (0.591-0.729)	0.679 (0.607-0.744)	0.695 (0.629-0.761)	0.998 (0.998-0.999)	0.998 (0.998-0.999)	0.999 (0.995-1.00)
Sadness	0.699[c] (0.645-0.748)	0.718[d] (0.664-0.766)	0.856[c,d] (0.818-0.895)	0.997 (0.997-0.998)	0.997 (0.997-0.997)	0.997 (0.991-1.00)

[a]*P* value corrected after Bonferroni *P*<.0021 (*P*=alpha/N).

[b]Significant difference between Linguistic Inquiry and Word Count 2001 and Linguistic Inquiry and Word Count 2007 at *P*<.0021.

[c]Significant difference between Linguistic Inquiry and Word Count 2001 and Linguistic Inquiry and Word Count 2015 at *P*<.0021.

[d]Significant difference between Linguistic Inquiry and Word Count 2007 and Linguistic Inquiry and Word Count 2015 at *P*<.0021.

Specificity

Specificity measured the proportion of nonemotional words that were accurately coded by LIWC 2001, LIWC 2007, or LIWC 2015 as not being indicative of emotion. Specificity for LIWC 2001, LIWC 2007, and LIWC 2015 was exceptional in all emotion categories (Table 4). Specificity values for LIWC 2001, LIWC 2007, and LIWC 2015 ranged from 0.955 for total emotional expression to 0.999 for anger. There were no differences in overall emotional expression, positive emotions, negative emotions, anxiety, anger, or sadness between LIWC 2001 and LIWC 2007.

PPV

PPV measured the probability that a word identified by LIWC 2001, LIWC 2007, and LIWC 2015 as being representative of emotional expression was in agreement with human rater coding of emotional expression. PPV for LIWC 2001, LIWC 2007, and LIWC 2015 was poor in all emotion categories (Table 5). The PPV values for all 3 versions ranged from 0.207 to 0.640. LIWC 2001's PPV was significantly better than LIWC 2007's and LIWC 2015's PPV in total emotion, negative emotion, and anxiety and was significantly better than LIWC 2007's PPV in positive emotion. Notably, there were no differences in anger or sadness.

Table 5. Linguistic Inquiry and Word Count 2001, 2007, and 2015 positive predictive value and negative predictive value with 95% CI (N=63)[a].

Emotion categories	2001 PPV[b] (95% CI)	2007 PPV (95% CI)	2015 PPV (95% CI)	2001 NPV[c] (95% CI)	2007 NPV (95% CI)	2015 NPV (95% CI)
Total affect	0.326[d,e] (0.315-0.336)	0.268[d] (0.259-0.277)	0.270[e] (0.254-0.286)	0.997 (0.997-0.997)	0.997 (0.997-0.998)	0.998 (0.996-0.999)
Positive Emotion	0.256[d] (0.244-0.268)	0.207[d] (0.197-0.217)	0.211 (0.191-0.231)	0.998 (0.998-0.998)	0.999 (0.999-0.999)	0.999 (0.997-1.00)
Negative Emotion	0.498[d,e] (0.479-0.516)	0.377[d] (0.361-0.395)	0.373[e] (0.348-0.398)	0.998 (0.998-0.998)	0.998 (0.998-0.998)	0.998 (0.996-1.00)
Anxiety	0.640[d] (0.605-0.675)	0.477[d] (0.446-0.508)	0.496 (0.454-0.538)	0.999 (0.999-0.999)	0.999 (0.999-0.999)	0.999 (0.996-1.00)
Anger	0.357[e] (0.307-0.409)	0.317 (0.273-0.366)	0.375[e] (0.306-0.444)	0.999 (0.999-0.999)	0.999 (0.999-0.999)	0.999 (0.995-1.00)
Sadness	0.389 (0.349-0.431)	0.351 (0.315-0.389)	0.377 (0.324-0.430)	0.999 (0.999-0.999)	0.999 (0.999-0.999)	0.999 (0.996-1.00)

[a]*P* value corrected after Bonferroni *P*<.0021 (*P*=alpha/N).

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]Significant difference between Linguistic Inquiry and Word Count 2001 and Linguistic Inquiry and Word Count 2007 at *P*<.0021.

[e]Significant difference between Linguistic Inquiry and Word Count 2001 and Linguistic Inquiry and Word Count 2015 at *P*<.0021.

NPV

NPVs measured the probability that a word not identified as emotion by LIWC 2001, LIWC 2007, and LIWC 2015 agreed with raters' judgment that the word was not associated with emotional expression. LIWC 2001, LIWC 2007, and LIWC 2015 have excellent NPVs across all emotion categories (Table 5). The NPV for LIWC 2001, LIWC 2007, and LIWC 2015 ranged from 0.997 for total emotional expression to 0.999 for anxiety, anger, and sadness. There were no significant differences between LIWC 2001, LIWC 2007, and LIWC 2015 with regard to NPV.

F Score

The results of the *F* score were compared using a test of difference, which was conducted to determine whether the difference between two proportions was significant, and it revealed that LIWC 2001 was significantly more precise in its evaluation of total emotional expression, positive emotion, and anxiety in comparison with LIWC 2007. Additionally, LIWC 2001 showed significantly more precision in its evaluation of total affect, positive emotion, negative emotion, and anxiety in comparison with LIWC 2015. Notably, there were no significant differences in accuracy between LIWC 2007 and LIWC 2015 (Table 6).

Table 6. Linguistic Inquiry and Word Count 2001, 2007, and 2015 F score values.

Emotion categories	2001 *F* score	2007 *F* score	2015 *F* score	Difference 2001-2007	Difference 2001-2015	Difference 2007-2015
Total affect	0.472	0.413	0.415	<.0001[a]	<.0001[a]	.875
Positive emotion	0.396	0.337	0.344	.0007[a]	.003[a]	.683
Negative emotion	0.561	0.516	0.511	.014	.007[a]	.786
Anxiety	0.735	0.622	0.635	<.0001[a]	.0003[a]	.654
Anger	0.464	0.433	0.487	.542	.654	.291
Sadness	0.499	0.497	0.523	.497	.544	.198

[a]*P* value corrected after Bonferroni *P*<.0083 (*P*=alpha/N).

Discussion

Successive versions of LIWC have become increasingly sensitive to identifying emotional expression. Our hypothesis that LIWC 2015 would significantly differ from LIWC 2001 and 2007 and be more efficacious in emotional identification was not supported. LIWC 2007 and 2015 were able to increase the previously established strength of LIWC 2001 in the identification of overall emotional expression emotion and positive emotion, whereas LIWC 2015 also increases the strength in the identification of sadness. However, LIWC 2007 and 2015 both exacerbate the existing weakness of LIWC 2001 in that many of the words it identifies as emotion do not match the ratings of human coders. Regarding the identification of nonemotional words, there was no improvement by LIWC 2015

or LIWC 2007. These findings suggest that although LIWC 2015 and LIWC 2007 had higher levels of emotional identification, more words were also inaccurately classified as emotion, compared with LIWC 2001. Therefore, although LIWC 2001, LIWC 2007, and LIWC 2015 measure a number of domains other than emotional expression, our findings suggest that all 3 versions, LIWC 2001, LIWC 2007, and LIWC 2015, have excellent sensitivity for detecting emotional expression, but LIWC 2001 has more precision with respect to PPV—the words it identifies as representing emotion are more likely than LIWC 2007 and LIWC 2015 to be in agreement with human raters.

The sensitivity levels for all 3 versions of LIWC indicate strength with regard to the identification of emotional content, such that they were highly sensitive to the identification of emotional expression. This is exceptionally important when analyzing text data where the expected prevalence of emotional expression is low or where the risk of overidentification is much lower than the risk of under-identification (eg, in evaluating suicide risk), where a missed instance could have grave consequences. There are many applications for which sensitivity, rather than accuracy, may be preferable, and LIWC 2007 and LIWC 2015 demonstrate an improved performance in sensitivity relative to LIWC 2001. However, there are potential consequences to the overidentification of emotional expression, particularly when accurate emotion recognition is required for a specific utterance.

In contrast to sensitivity, the PPV was fairly poor for LIWC 2001, LIWC 2007, and LIWC 2015. LIWC 2001 produced a significantly more precise performance with regard to PPV compared with LIWC 2007 and LIWC 2015, whereas there were no significant differences between LIWC 2007 and LIWC 2015. Evaluation of the F score, which balances both PPV and sensitivity, revealed that LIWC 2001 was superior to LIWC 2007 in the emotional identification of overall affect, positive emotions, and anxiety. The remaining categories were not significantly different, indicating that LIWC 2001 and LIWC 2007 performed similarly in their identification of those emotion categories (eg, negative emotion, anger, and sadness). Additionally, with regard to *F* scores, LIWC 2001 was superior to LIWC 2015 in terms of total overall affect, positive emotions, negative emotions, and anxiety. The remaining categories were not significantly different, indicating that LIWC 2001 and LIWC 2015 performed similarly in their identification of those emotion categories (eg, anger and sadness). Finally, there were no significant differences in the F scores of emotional identification between LIWC 2007 and LIWC 2015 across the emotion categories (eg, overall affect, positive emotions, negative emotions, anxiety, anger, and sadness). These results indicate that LIWC 2001 is more inclined to accurately identify emotion in accordance with human coders when compared with LIWC 2007 and LIWC 2015, wherein one balances the problem of overidentification against the problem of under-identification, which is pertinent when considering at-risk populations (eg, suicidal patients). Considering that human coders are the gold standard for emotional identification and that LIWC 2001 provides results that are most similar to those of human coders, LIWC 2001 is superior to LIWC 2007 and LIWC 2015.

As previously mentioned, Pennebaker et al [14] made a number of changes to the LIWC 2007 and LIWC 2015 dictionaries, which may have resulted in the decrease in precision of the subsequent versions. Although the alterations to LIWC 2007 and LIWC 2015 resulted in improvements in sensitivity, these changes did not improve LIWC 2001's previously established flaws. The alterations to the LIWC 2007 and LIWC 2015 dictionary may have resulted in the increased identification but a decrease in the precision of the identification.

LIWC 2001 may present as superior in its emotional identification over LIWC 2007 and LIWC 2015, yet the accuracy in its performance is highly dependent upon the population being evaluated. The PPV is dependent upon the prevalence in the population, meaning it can vary based on the sample utilized, whereas sensitivity may remain the same despite what population is being evaluated [29]. More specifically, cancer patients have been found to express more emotion than a physically healthy population [30], meaning that the prevalence of expressed emotion is higher for the sample utilized in this study than that of the general population. Considering prevalence rates or emotional expression in the cancer population, LIWC 2001 and LIWC 2007 are likely to produce poorer PPVs if being utilized with the emotional expression of a nonclinical population. Text that is likely to have nuanced emotion, such as is often the case in the experience of cancer, will also be less likely to be related to accuracy when coded by LIWC. LIWC 2001 and LIWC 2007 currently have a high rate of false positives, which may increase when evaluating a less emotionally expressive population or decrease when evaluating a more emotionally expressive population. Ultimately, the LIWC programs would benefit from further validation utilizing alternative populations with varying levels of emotional expression.

The initial validation of LIWC 2001, 2007, and 2015 relied heavily on simple correlations between LIWC codes and judge's ratings of emotional content. Correlation analyses describe the extent to which the variables covary, but not the accuracy of identification. For this, measures of testing accuracy are more appropriate. Conducting analyses such as a test of proportions allows users to see the weaknesses and strengths of the relationship and what factors contribute to the strengths of that relationship. Additionally, future studies may also benefit from evaluating writers' intentions related to emotional expression to further validate both LIWC and human coders' ratings. For example, the LIWC classification results (eg, negative emotion) could have been reviewed while also conducting a review with the text writers, ensuring that they intended to express negative emotions, thus confirming the LIWC classification. A machine learning approach, in which rules for coding are created as the raw data are under analysis, in combination with LIWC is another possibility for increasing the overall accuracy as well as making the coding procedure more sample dependent. With a machine learning approach, automated coding rules or algorithms are generated from patterns seen in other coding methods, such as manual coding, potentially allowing for increased accuracy without the time-consuming nature of manual coding [13]. Emotions are multifaceted, making them much more difficult to accurately identify when simplified to one

modality. It is important to note that there are many other ways to code emotion in text, and this study focused on validating a program that is commonly used in the field of web-based behavioral sciences.

It must be noted that there are some limitations to this study. The narratives utilized in this study were obtained from women diagnosed with breast cancer. Research has indicated that female cancer patients express more emotion than male counterparts [30]. Additionally, cancer patients are more inclined to endorse affective disorders, such as anxiety, which may impact their emotional expression [31]. Additionally, based on the specific circumstances these women faced (eg, cancer diagnosis, treatment, and outcomes), this could have limited the range of emotions that may have been discussed compared with a healthy population. On the basis of the population utilized, results may be limited to cancer survivors rather than the general population. Furthermore, there were few emotions evaluated (eg, overall affect, positive emotions, negative emotions, anger, anxiety, and sadness), which does not reflect the full range of emotions experienced. This limited range of emotions measured may not accurately reflect the emotions expressed (eg, frustration, excitement, and fear). Finally, the sample utilized was highly homogenous with respect to gender, ethnicity, level of education, and even health status, further limiting the generalizability of the findings.

In contrast to the sample, additional limitations present within this study include intricacies specific to LIWC. As previously mentioned, the classification of the word "like" has been altered across each iteration of LIWC. Most notably, when utilizing LIWC 2015 in text analysis, it may be most accurate to manually evaluate each instance of the word "like" to determine the contextual meaning of the word to ensure appropriate classification of each word. Despite this, manual evaluation determining the context of the word "like" was not performed before conducting automatic analysis with LIWC 2015, as this defeats the utility and efficiency that automatic programs of text analysis offer.

Although human coders are the gold standard for emotional identification in text data, due to the time and cost associated with evaluating such large volumes of data, human coding is not always practical. In addition, manual coding of text, although considered the gold standard, is not entirely accurate. There is inevitably some ambiguity in any attempt to capture the emotion expressed by another person. Text-based data also leave absent, nonverbal expression, leaving fewer cues to code. On the basis of the results, LIWC 2001 is clearly superior to LIWC 2007 and LIWC 2015 with respect to overall precision, but LIWC 2007 and LIWC 2015 are more sensitive to identifying overall expressed emotions. The PPV is highly dependent on the prevalence of emotion in the specific population, such that the more emotion presented in a population, the more accurate the analysis will likely be. Considering the high prevalence of emotion in a cancer population and that LIWC 2001 performed significantly better than LIWC 2007 and LIWC 2015, this indicates that for a population with much less emotional expression, LIWC 2007 and LIWC 2015 will still perform significantly poorer than LIWC 2001. LIWC 2001 seems to have good validity in emotional identification and presents as a reasonable tool for identification of emotion in text data, which is important in the increasingly digital world.

Acknowledgments

The authors have no relevant affiliations or financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This manuscript is not under review elsewhere, has not been published previously, or accepted for publication with another entity. The Loma Linda University ethics committee declared this study exempt from review. No writing assistance was utilized in the production of this manuscript. A portion of this project was funded by an intradepartment graduate student grant from Loma Linda University.

Conflicts of Interest

None declared.

References

1. Baddeley JL, Daniel GR, Pennebaker JW. How Henry Hellyer's use of language foretold his suicide. Crisis 2011;32(5):288-292. [doi: 10.1027/0227-5910/a000092] [Medline: 21940249]
2. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 2001;63(4):517-522. [doi: 10.1097/00006842-200107000-00001] [Medline: 11485104]
3. Ludwig S, de Ruyter K, Friedman M, Brüggen E, Wetzels M, Pfann G. More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. J Mark 2013 Jan;77(1):87-103. [doi: 10.1509/jm.11.0560]
4. Peden BF, Carroll DW. Ways of writing: linguistic analysis of self-assessment and traditional assignments. Teach Psycol 2008;35:313-318. [doi: 10.1080/00986280802374419]
5. Bhatia MS, Verma SK, Murty OP. Suicide notes: psychological and clinical profile. Int J Psychiatry Med 2006;36(2):163-170. [doi: 10.2190/5690-CMGX-6A1C-Q28H] [Medline: 17154146]
6. Chochinov HM, Wilson KG, Enns M, Lander S. Depression, hopelessness, and suicidal ideation in the terminally ill. Psychosomatics 1998;39(4):366-370. [doi: 10.1016/S0033-3182(98)71325-8] [Medline: 9691706]

7. Schaefer ZK, Korkmaz G. Human vs Automated Text Analysis: Estimating Positive and Negative Affect. In: Proceedings of the 27th ACM Conference on Hypertext and Social Media. 2016 Presented at: HT'16; July 10-13, 2016; Halifax, Nova Scotia. [doi: 10.1145/2914586.2914634]
8. Ziemer KS, Korkmaz G. Using text to predict psychological and physical health: a comparison of human raters and computerized text analysis. Comput Hum Behav 2017 Nov;76:122-127. [doi: 10.1016/j.chb.2017.06.038]
9. Landless B, Walker MS, Kaimal G. Using human and computer-based text analysis of clinical notes to understand military service members' experiences with therapeutic writing. Arts Psychother 2019 Feb;62:77-84 [FREE Full text] [doi: 10.1016/j.aip.2018.10.002]
10. Kidd SA. The role of qualitative research in psychological journals. Psychol Methods 2002 Mar;7(1):126-138. [doi: 10.1037/1082-989x.7.1.126] [Medline: 11928887]
11. Pennebaker J, Boothe RE, Francis ME. Linguistic Inquiry and Word Count. Austin, TX: Psychology Press; 2001.
12. Bantum EO, Owen JE. Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. Psychol Assess 2009 Mar;21(1):79-88. [doi: 10.1037/a0014643] [Medline: 19290768]
13. Bantum EO, Elhadad N, Owen JE, Zhang S, Golant M, Buzaglo J, et al. Machine learning for identifying emotional expression in text: improving the accuracy of established methods. J Technol Behav Sci 2017 Apr 4;2(1):21-27. [doi: 10.1007/s41347-017-0015-5] [Medline: 32885036]
14. Pennebaker J, Chung CK, Ireland M, Gonzalez A, Booth R. The Development and Psychometric Properties of LIWC. Repository Home. 2007. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2020-09-21]
15. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related Twitter posts. Crisis 2017 Sep;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]
16. Sonnenschein AR, Hofmann SG, Ziegelmayer T, Lutz W. Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. Cogn Behav Ther 2018 Jul;47(4):315-327. [doi: 10.1080/16506073.2017.1419505] [Medline: 29345528]
17. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPsych'15; June 5, 2015; Denver, Colorado. [doi: 10.3115/v1/w15-1201]
18. Lyons M, Aksayli ND, Brewer G. Mental distress and language use: linguistic analysis of discussion forum posts. Comput Hum Behav 2018 Oct;87:207-211. [doi: 10.1016/j.chb.2018.05.035]
19. Westgate CL, Shiner B, Thompson P, Watts BV. Evaluation of veterans' suicide risk with the use of linguistic detection methods. Psychiatr Serv 2015 Oct;66(10):1051-1056. [doi: 10.1176/appi.ps.201400283] [Medline: 26073409]
20. Kwang T, Crockett EE, Sanchez DT, Swann WB. Men seek social standing, women seek companionship: sex differences in deriving self-worth from relationships. Psychol Sci 2013 Jul 1;24(7):1142-1150. [doi: 10.1177/0956797612467466] [Medline: 23658253]
21. Tumasjan A, Sprenger T. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In: Proceedings of the Fourth International Conference on Weblogs and Social Media. 2010 Presented at: ICWSM'10; May 23-26, 2010; Washington, DC, USA URL: https://www.researchgate.net/publication/215776042_Predicting_Elections_with_Twitter_What_140_Characters_Reveal_about_Political_Sentiment
22. Ireland ME, Slatcher RB, Eastwick PW, Scissors LE, Finkel EJ, Pennebaker JW. Language style matching predicts relationship initiation and stability. Psychol Sci 2011 Jan;22(1):39-44. [doi: 10.1177/0956797610392928] [Medline: 21149854]
23. Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol 1998 Feb;66(1):7-18. [doi: 10.1037//0022-006x.66.1.7] [Medline: 9489259]
24. Weisz JR, Hawley KM. Finding, evaluating, refining, and applying empirically supported treatments for children and adolescents. J Clin Child Psychol 1998 Jun;27(2):206-216. [doi: 10.1207/s15374424jccp2702_7] [Medline: 9648037]
25. Owen J, Klapow JC, Roth DL, Shuster JL, Bellis J, Meredith R, et al. Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer. Ann Behav Med 2005 Aug;30(1):54-64. [doi: 10.1207/s15324796abm3001_7] [Medline: 16097906]
26. Owen JE, Giese-Davis J, Cordova M, Kronenwetter C, Golant M, Spiegel D. Self-report and linguistic indicators of emotional expression in narratives as predictors of adjustment to cancer. J Behav Med 2006 Aug;29(4):335-345. [doi: 10.1007/s10865-006-9061-8] [Medline: 16845583]
27. Pennebaker J, Boyd TL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC. Repository Home. 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2020-09-21]
28. Faul F, Erdfelder E, Lang A, Buchner A. G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
29. Altman DG, Bland JM. Diagnostic tests 2: predictive values. Br Med J 1994 Jul 9;309(6947):102 [FREE Full text] [doi: 10.1136/bmj.309.6947.102] [Medline: 8038641]

30. Linden W, Vodermaier A, Mackenzie R, Greig D. Anxiety and depression after cancer diagnosis: prevalence rates by cancer type, gender, and age. J Affect Disord 2012 Dec 10;141(2-3):343-351. [doi: 10.1016/j.jad.2012.03.025] [Medline: 22727334]
31. Mitchell AJ, Ferguson DW, Gill J, Paul J, Symonds P. Depression and anxiety in long-term cancer survivors compared with spouses and healthy controls: a systematic review and meta-analysis. Lancet Oncol 2013 Jul;14(8):721-732. [doi: 10.1016/s1470-2045(13)70244-4] [Medline: 23759376]

Abbreviations

LIWC: Linguistic Inquiry and Word Count
NPV: negative predictive value
PPV: positive predictive value

Edited by C Lovis, G Eysenbach; submitted 13.02.20; peer-reviewed by R Boyd, J Zhu, J Groarke, S Mohammad; comments to author 28.04.20; revised version received 02.07.20; accepted 18.08.20; published 30.10.20.

Please cite as:
McDonnell M, Owen JE, Bantum EO
Identification of Emotional Expression With Cancer Survivors: Validation of Linguistic Inquiry and Word Count
JMIR Form Res 2020;4(10):e18246
URL: https://formative.jmir.org/2020/10/e18246
doi:10.2196/18246
PMID:33124986

Original Paper

Investigating the Impact of COVID-19 Lockdown on the Psychological Health of University Students and Their Attitudes Toward Mobile Mental Health Solutions: Two-Part Questionnaire Study

Nidal Drissi[1], MSc; Ayat Alhmoudi[1], BSc; Hana Al Nuaimi[1], BSc; Mahra Alkhyeli[1], BSc; Shaikha Alsalami[1], BSc; Sofia Ouhbi[1], PhD

United Arab Emirates University, Al Ain, United Arab Emirates

Corresponding Author:
Sofia Ouhbi, PhD
United Arab Emirates University
Al Ain
United Arab Emirates
Phone: 971 37135568
Email: sofia.ouhbi@uaeu.ac.ae

Abstract

Background: The COVID-19 outbreak was first reported to the World Health Organization on December 31, 2019, and it was officially declared a public health emergency of international concern on January 30, 2020. The COVID-19 outbreak and the safety measures taken to control it caused many psychological issues in populations worldwide, such as depression, anxiety, and stress.

Objective: The objectives of this study were to assess the psychological effects of the lockdown due to the COVID-19 outbreak on university students in the United Arab Emirates (UAE) and to investigate the students' awareness of mobile mental health care apps as well as their attitudes toward the use of these apps.

Methods: A two-part self-administered web-based questionnaire was delivered to students at United Arab Emirates University. The first part of the questionnaire assessed the mental state of the participants using the 12-item General Health Questionnaire (GHQ-12), while the second part contained questions investigating the participants' awareness of and attitudes toward mental health care apps. Students were invited to fill out the web-based questionnaire via social media and mailing lists.

Results: A total of 154 students participated in the survey, and the majority were female. The results of the GHQ-12 analysis showed that the students were experiencing psychological issues related to depression and anxiety as well as social dysfunction. The results also revealed a lack of awareness of mental health care apps and uncertainty regarding the use of such apps. Approximately one-third of the participants (44/154, 28.6%) suggested preferred functionalities and characteristics of mobile mental health care apps, such as affordable price, simple design, ease of use, web-based therapy, communication with others experiencing the same issues, and tracking of mental status.

Conclusions: Like many groups of people worldwide, university students in the UAE were psychologically affected by the lockdown due to the COVID-19 outbreak. Although apps can be useful tools for mental health care delivery, especially in circumstances such as those produced by the outbreak, the students in this study showed a lack of awareness of these apps and mixed attitudes toward them. Improving the digital health literacy of university students in the UAE by increasing their awareness of mental health care apps and the treatment methods and benefits of the apps, as well as involving students in the app creation process, may encourage students to use these tools for mental health care.

(JMIR Form Res 2020;4(10):e19876) doi:10.2196/19876

KEYWORDS

COVID-19; GHQ-12; mobile; apps; m-health; m-mental health; UAE; attitudes; university students; questionnaire

Introduction

On December 31, 2019, the World Health Organization (WHO) was informed of several cases of pneumonia of unknown cause detected in Wuhan City, China [1]. On January 7, 2020, the cause was reported to be a new coronavirus, which was then referred to as "2019-nCoV" [1]. By January 20, 2020, the virus had spread, and multiple cases were reported in four countries [1]. On January 30, 2020, the WHO declared the outbreak to be a public health emergency of international concern [2]. By February 11, 2020, the virus had spread to more than 24 countries in addition to China, and the coronavirus disease was officially named COVID-19 [3]. By May 2020, the virus had spread to all regions of the world [4]. More than 3 million cases and 248,847 deaths were reported worldwide as of May 4, 2020. To limit the spread and risk of the virus, the WHO advised people to practice social distancing and to stay at home [5]. Countries took different safety measures and precautions to prevent the spread of the disease. Several countries declared obligatory lockdowns, travel was halted and airports were shut down, and many work spaces, schools and universities were closed.

Lockdown-related stressors such as concern about the duration of lockdown, fear of infection, boredom, inadequate information [6], and fear of the unknown [7] had significant psychological effects on people, including posttraumatic stress disorder symptoms, anger, confusion, fear, worry, sadness, and elevated anxiety and stress [6,8]. As part of its application of precautions and safety measures to address the outbreak, the United Arab Emirates (UAE) closed its universities and stopped all related activities. To investigate the impact of the lockdown on university students in the UAE, we conducted a mental health assessment based on the 12-item General Health Questionnaire (GHQ-12). The GHQ-12 has been used in previous studies to assess the mental health of students and has shown positive reliability results. A study using the GHQ-12 to assess the psychological state of Malaysian college students affirmed that the GHQ-12 is a good tool for assessing the overall psychological well-being of students [9]. In another study, the questionnaire was used with Australian college students; the results suggested that the GHQ-12 is a viable tool with good clinical utility and that it can be implemented as part of routine school screening procedures to help identify young people at risk of depressive and anxiety disorders [10]. Another study used the GHQ-12 with Tehrani university students, and it was concluded that the questionnaire is suitable for screening psychopathology in university students [11].

Moreover, the validity of the psychometric properties of the GHQ-12 was asserted in several studies. The GHQ-12 was assessed against the Clinical Interview Schedule (CIS) in three primary care settings in Brazil and was found to be acceptable and valid [12]. The GHQ-12 was also assessed in a sample of dermatological patients against the Skindex-29, which is a tool to evaluate the impact of skin conditions on the quality of life of patients. The results showed that the GHQ-12 is a reliable and valid instrument [13]. Furthermore, the GHQ-12 was reported to be robust and suitable for use as a screening instrument in a study conducted by the WHO in which the 28-item General Health Questionnaire (GHQ-28) was compared to the GHQ-12 [14].

Identifying existing psychological issues is an important step; however, it is more important to deliver mental health care when it is needed. This delivery is currently challenging because of the safety measures being applied to prevent the spread of COVID-19. Telehealth is a health care delivery method that, like tele-education and telework, currently seems to be the safest approach. Telehealth can be applied to mental health care via several methods, one of which is mobile apps. Apps are highly suitable to provide services while safety measures are being applied during the pandemic. They can be used to provide mental health care without need for human contact, and the user can benefit from care delivered via the app without needing to leave their home and risk exposure to the virus. Apps can also help overcome several pre-existing barriers to mental health care delivery in addition to those created by the COVID-19 outbreak, such as cost problems, stigma, and distance or shortage of mental health professionals. Apps have also shown promising results in the management of many mental issues, such as anxiety, depression, and stress [15,16].

This study had two main goals. The first goal was to assess the mental state of university students during the first period of the lockdown caused by the COVID-19 outbreak. The second goal of the study was to investigate the students' awareness of mobile apps for mental health care and their willingness to use them as well as to discover the features they would like a mental health care app to have and the factors that would encourage them to use such apps. A two-part web-based questionnaire was delivered to United Arab Emirates University (UAEU) students via social media and email during the first two weeks of the lockdown imposed in the UAE. The first part of the form consisted of the GHQ-12 questionnaire, and the second part of the form included questions investigating the students' attitudes toward mobile apps for mental health care. The study verified the following hypotheses: (h1) the lockdown has an impact on the psychological state of the university students in the UAE; (h2) university students are aware of the existence of mobile apps for mental health care; and (h3) university students have positive attitudes toward mobile mental health care apps and are open to the use of such solutions.

Methods

Research Design

This study constitutes two main parts. The objective of the first part was to assess the psychological health of university students. The objective of the second part was to investigate the students' awareness of and attitudes toward mHealth apps for mental health care.

Recruitment and Data Collection

A self-administered web-based questionnaire was created using Google Forms and sent to UAEU students via social media and mailing lists. Recruiting participants through social media has been found to be effective and time-efficient [17]. The included participants were a self-selecting sample, as participation in the survey was voluntary and participants were not offered any

incentives. The questionnaire consisted of 20 questions and was made available on the internet for two weeks, from March 15, 2020, to March 29, 2020. The questions included 13 predefined multiple-choice questions, 4 yes/no questions, and 3 open questions. None of the questions were personal questions that could reveal the participants' identity. The questionnaire was tested by the authors before it was sent to students. The estimated time to complete the questionnaire was 4 minutes, which was stated in the questionnaire. The questionnaire also stated that the responses would be anonymous. Prior to data collection, permission was obtained from the relevant authorities at UAEU, and a psychologist was consulted to determine the appropriateness of the questionnaire for the target respondents. The information provided about the questionnaire was based on the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [18].

Survey Questions

The web-based questionnaire contained 20 questions. The first three questions aimed to obtain basic information and characteristics of the participants (age, gender, and academic major). The remaining questions were divided into two parts.

Part 1: GHQ-12

In this part, we used the GHQ-12 [19] to measure the psychological health of the survey participants. The GHQ-12 can be analyzed as a single dimension psychological health test [20]. However, many researchers have suggested that it can be divided into two or three specific and meaningful factors, in which each factor is composed of several items from the questionnaire. Gribbin and Worsley [21] proposed a three-factor approach: anxiety/depression, social dysfunction, and loss of confidence. Andrich and van Schoubroeck [22] suggested that positively worded items or questions constitute one factor and negatively worded ones constitute another. Politi et al [23] identified two factors: general dysphoria and social dysfunction. Martin [24] proposed three factors: self-esteem, stress, and successful coping. When compared to other methods and applied to different samples, the three-factor model proposed by Gribbin and Worsley [21], including anxiety and depression (4 items), social dysfunction (6 items), and loss of confidence (2 items), was found to give the best fit [25,26]. Therefore, we used this model in this study.

Table 1 presents the 12 items of the GHQ-12. Table 2 presents the association between the GHQ-12 items and the three psychological factors. A study by Gao et al [27] showed that the loadings of the items on their associated factors are very close, ranging from 0.72 to 0.9. The correlation between the three factors was also found to be very high, ranging from 0.83 to 0.9 [27]. Based on these results, and to simplify the assessment, we assumed that all items had the same loadings on their associated factors.

The potential answers to the GHQ-12 items range from "better/healthier than normal" through "same as usual" and "worse/more than usual" to "much worse/more than usual." The answers reflect the difference between the psychological health states of the participants when they responded to the questionnaire and when they considered their health states to be normal. The answers can be scored in four different ways: GHQ scoring (0-0-1-1), C-GHQ scoring ((0-1-1-1) for negative items and (0-0-1-1) for positive items), Likert scoring (0-1-2-3), and modified Likert scoring (0-0-1-2).

This study investigates the severity of the identified psychological issues in addition to their prevalence; therefore, we used the Likert scoring method to score the answers because it produces a wider and smoother score distribution, which helps to assess severity [28]. Each item of the GHQ-12 had 4 possible answers that were scored from 0 to 3. A higher score indicated a more severe condition. The scores for each answer of the 12 items are presented in Table 1.

Table 1. Questions, answers, and scores of the 12-item General Health Questionnaire.

Item ID	Question	Answers	Score
1	Have you recently been able to concentrate on what you are doing?	Better than usual	0
		Same as usual	1
		Less than usual	2
		Much less than usual	3
2	Have you recently lost much sleep over worry?	Not at all	0
		No more than usual	1
		Rather more than usual	2
		Much more than usual	3
3	Have you recently felt you were playing a useful part in things?	More so than usual	0
		Same as usual	1
		Less than usual	2
		Much less than usual	3
4	Have you recently felt capable of making decisions about things?	More so than usual	0
		Same as usual	1
		Less than usual	2
		Much less than usual	3
5	Have you recently felt constantly under strain?	Not at all	0
		No more than usual	1
		Rather more than usual	2
		Much more than usual	3
6	Have you recently felt you couldn’t overcome your difficulties?	Not at all	0
		No more than usual	1
		Rather more than usual	2
		Much more than usual	3
7	Have you recently been able to enjoy your normal day-to-day activities?	More so than usual	0
		Same as usual	1
		Less so than usual	2
		Much less than usual	3
8	Have you recently been able to face up to your problems?	More so than usual	0
		Same as usual	1
		Less so than usual	2
		Much less than usual	3
9	Have you recently been feeling unhappy and depressed?	Not at all	0
		No more than usual	1
		Rather more than usual	2
		Much more than usual	3
10	Have you recently been losing confidence in yourself?	Not at all	0
		No more than usual	1
		Rather more than usual	2
		Much more than usual	3
11	Have you recently been thinking of yourself as a worthless person?	Not at all	0
		No more than usual	1

Item ID	Question	Answers	Score
		Rather more than usual	2
		Much more than usual	3
12	Have you recently been feeling reasonably happy, all things considered?	More so than usual	0
		About same as usual	1
		Less so than usual	2
		Much less than usual	3

Table 2. Associations between items and psychological factors of the 12-item General Health Questionnaire.

Psychological factor	Associated item IDs
Anxiety and depression	2, 5, 6, and 9
Social dysfunction	1, 3, 4, 7, 8, and 12
Loss of confidence	10 and 11

The answers of the participants were collected and scored. Based on the maximum score of each psychological factor and the total score, the answers were categorized to represent three severity categories: normal, high, and severe.

For the total score of the GHQ-12, the normal state score category ranged from 0 to 12, the high-risk score category ranged from 13 to 24, and the severe case score category ranged from 25 to 36. For the anxiety and depression factor, the normal state scores ranged from 0 to 4, the high-risk category scores ranged from 5 to 8, and the severe case scores ranged from 9 to 12. For the social dysfunction factor, the normal state scores ranged from 0 to 6, the high-risk scores ranged from 7 to 12, and the severe case scores ranged from 13 to 18. For the "loss of confidence" factor, the normal state scores ranged from 0 to 2, the high risk state included scores of 3 and 4, and the severe case state included scores of 5 and 6.

Part 2: Awareness of and Attitudes Toward Mobile Mental Health Apps

The second part of the questionnaire served to investigate the university students' knowledge of mental health apps and their opinions about the use of these apps for mental health care. The questions were formulated based on questions retrieved from related literature investigating similar topics [29-32]. As shown in Table 3, the questions were formulated mainly to determine if the students had any previous knowledge or experience with mental health apps; understand their willingness to consider using a mental health app; and understand the preferences and factors that could encourage them to use apps for mental health care.

Questions Q1 and Q2 were yes/no questions. Questions Q3 and Q4 had possible answers of "Yes," "No," or "I don't know." If a participant answered "Yes" to Q4, they were asked to provide the reason for their choice in Q4.1, in which the participant could choose from a predefined list of reasons, including cost problems, stigma related to mental problems, distance from mental health care professionals, shortage of mental health care professionals, and lack of knowledge (information) on mental health; they could also choose the "Other" option and express their own reasons. Q5 was an open question.

Table 3. Questions in Part 2 of the questionnaire related to mental health care apps.

Item ID	Question
1	Have you ever heard of mobile mental health apps?
2	Have you ever used a mobile app for your mental well-being?
3	Would you be open to using a mobile app for your mental well-being in the future?
4	Would you prefer using a mobile app over consulting with a mental health care specialist?
4.1	If yes, why?
5	What would you like to see in an app for mental health care?

Correlations

The correlations between the age and the total GHQ-12 score as well as between the gender and the total GHQ-12 score were calculated using the Pearson correlation coefficient, which has a value between +1 and –1. A value of +1 indicates a perfect positive correlation between the sets of data, while a value of –1 indicates a perfect negative correlation. A value of 0 indicates no correlation between the sets; the closer the correlation value is to 0, the weaker the correlation is.

Summary of the Analytical Strategy

For the GHQ-12 items, the participants answered each GHQ-12 item with "better/healthier than normal," "same as usual,"

"worse/more than usual," or "much worse/more than usual." The collected answers were converted to numerical scores based on the Likert scoring (0-1-2-3) method. The scores were regrouped to investigate the psychosocial factors presented in Table 2 (anxiety and depression, social dysfunction, and loss of confidence). A total GHQ-12 score was calculated based on the scores of all the items. The total score and scores of the psychological factors were categorized into three categories: normal state, high-risk state, and severe state. The higher the score, the more severe the participant's condition.

For the questions related to mental health care apps, the participants answered questions Q1 to Q4 using predefined options. The answers to questions Q1 to Q4 were then quantified and analyzed. For the open questions (Q5 and the "Other "option in Q4.1), the participants provided short answers. The answers of these open questions were then categorized based on their meanings and presented in a list.

Results

Sample Characteristics

A total of 154 university students participated in the survey; the majority of the participants were female (113, 73.4%). The academic majors of the majority of the respondents fell into the categories of science, mathematics, and technology (89/154, 57.8%). Table 4 summarizes the main characteristics of the participants.

Table 4. Characteristics of the study participants (N=154).

Variable	Value
Gender, n (%)	
Female	113 (73.4)
Male	41 (26.6)
Age (years)	
15-22, n (%)	118 (76.6)
≥23, n (%)	36 (23.4)
Mean (SD)	22.45 (5.87)
Academic major, n (%)	
Science, mathematics, and technology	89 (57.8)
Business	24 (15.6)
Health and medicine	10 (6.5)
Social sciences	5 (3.2)
Public and social services	5 (3.2)
Other	21 (13.6)

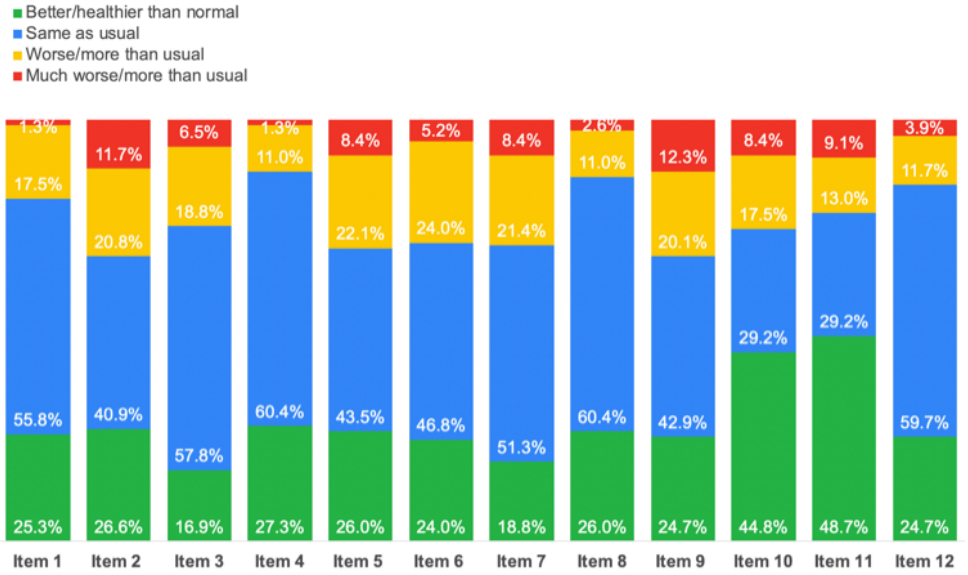
Part 1 Results

Results by Item

The results of the analysis of the items identified specific issues that the students suffered from more during the period of the survey. If the participant's answer to an item indicated a better/healthier state than usual or the same state as usual, the issue investigated by the item was not considered to be more present than usual for the participant; however, if the answer indicated a worse or much worse state than usual, it reflected an abnormal presence of the issue during the survey period. Figure 1 presents the answers to the GHQ-12 items.

One third of the 154 participants (50, 32.5%) had sleep issues (Item 2): 32 (20.8%) reported worse sleep than usual and 18 (11.7%) reported much worse sleep than usual. The same number of participants (50/154, 32.5%) showed elevated states of unhappiness and depression (Item 9), with 31/154 (20.1%) answering worse than usual and 19/154 (12.3%) answering much worse than usual. One-third of the participants (47/154, 30.5%) reported feeling constantly under strain (Item 5), with 34/154 (22.1%) answering worse than usual and 8.4% (13/154) answering much worse than usual. The issues for which participants showed the least elevated states were the capabilities of making decisions and facing problems (Items 4 and 8, respectively); 19/154 participants (12.3%) reported worse capability of making decisions, and 21/154 (13.6%) reported worse ability to face problems.

Figure 1. Proportions of participants' answers to the 12-item General Health Questionnaire.



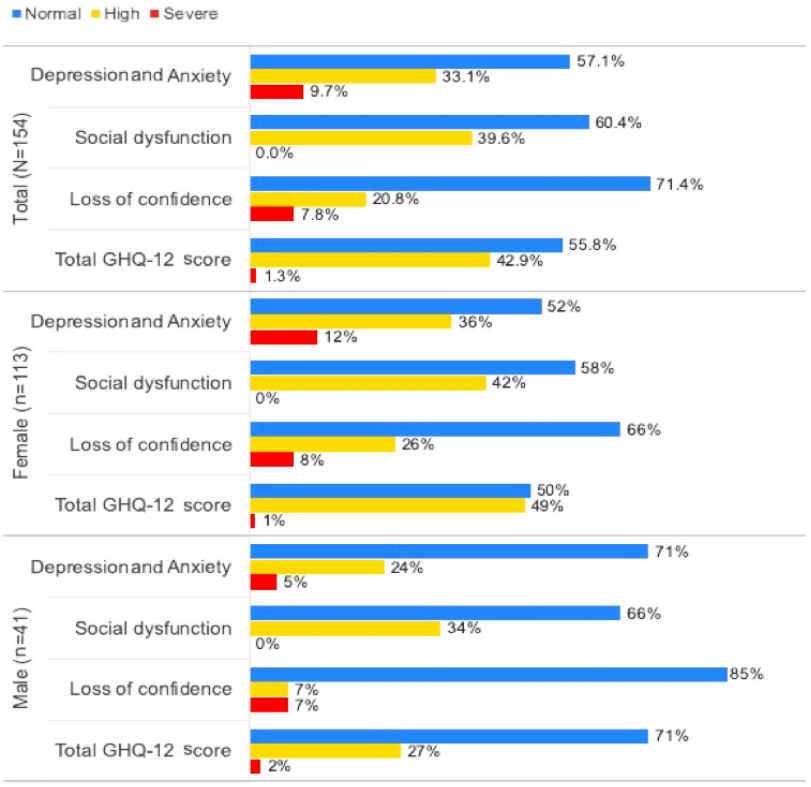
Total Score and Psychological Factors

Analysis of the total scores and the scores of each psychological factor gave a general idea of the students' mental states and the mental issues they were more susceptible to. Figure 2 presents the severity categories of each psychological factor as well as the total score of the GHQ-12. The figure also presents a classification by gender for each category.

More than one-third of the participants (66/154, 42.9%) had high scores indicating high risk of having mental issues, and a small number of participants (2/154, 1.3%) had very high scores and were classified as severe cases. The issues of the participants mainly seemed to relate to depression and anxiety.

Weak correlations were found between the age of the participants and the total score of the GHQ-12 (–0.101) and between the gender of the participants and the total score of the GHQ-12 (–0.128).

Figure 2. Classification of GHQ-12 answers by severity category and gender. GHQ-12: 12-item General Health Questionnaire.



Part 2 Results

Answers to the Questions

Figure 3 presents the responses to questions Q1 to Q4 related to mental health care apps.

Half of the participants (78/154, 50.6%) had never heard of mental health care apps (Q1). The majority of the participants (114/154, 74.0%) had never used an app for mental health care (Q2). More than half of the participants (82/154, 53.2%) showed uncertainty regarding their willingness to use mental health care apps in the future (Q3). Of the 154 participants, 83 (53.9%) justified their answers of "Yes" or "I don't know" to Q4 by the reasons shown in Figure 4.

Figure 3. Answers to questions Q1, Q2, Q3, and Q4 pertaining to mental health apps.

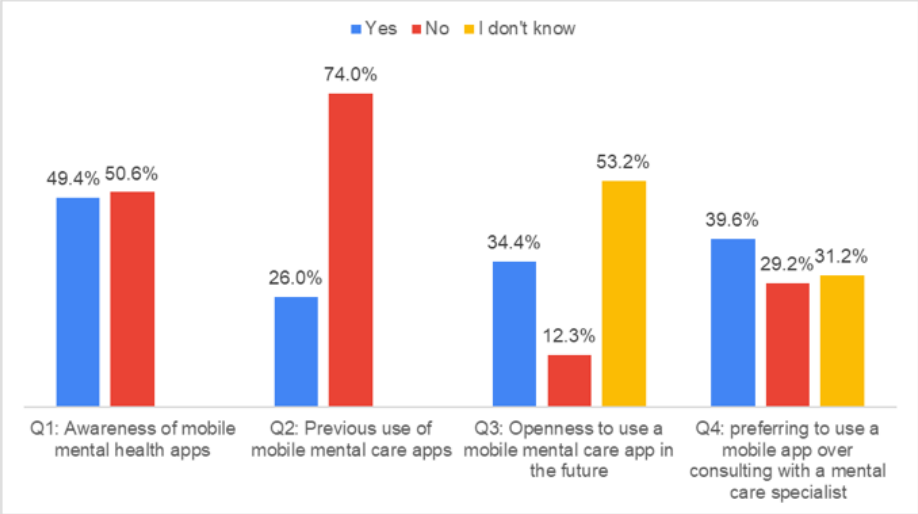
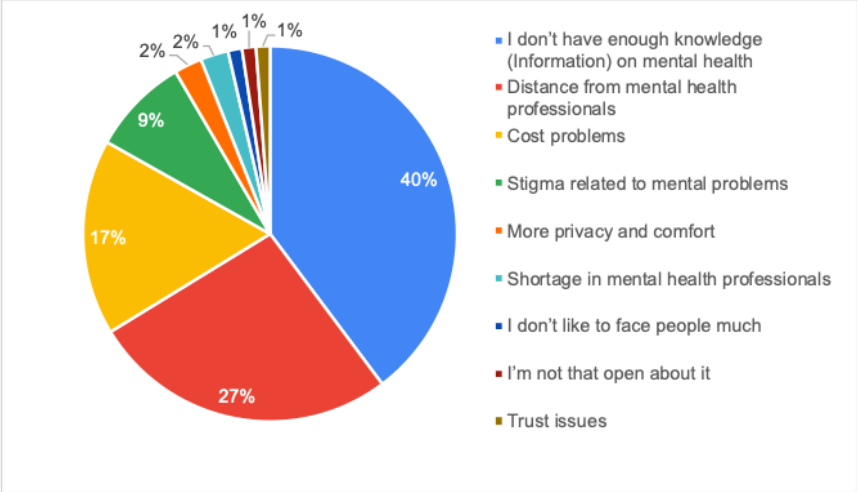


Figure 4. Answers to Q4.1 indicating the participants' reasons for preferring the use of a mobile app over consulting with a mental health care specialist (n=83).



To identify the features that the students would need or prefer to see in mental health care apps, the participants were asked an open question (Q5) about what would they like to see in an app for mental health care. Almost one-third of the participants (44/154, 28.6%) answered the open question Q5 with the functionality and characteristics that they thought mental health care apps should have. Their answers are grouped and presented in Textbox 1.

Textbox 1. Preferred functionality and characteristics of mental health care apps indicated by the survey participants.

Functionality

- Web-based therapy
- Communication with others experiencing same issues
- Tracking mental status
- Advice from specialists
- Motivational statements
- Educational content
- Games
- Recommendations of activities and tips
- General health management
- Emergency features
- Tracking of progress
- Stories

Characteristics

- Anonymity
- Simple design and ease of use
- Affordable price
- Diversity of features

Associations Between the Answers

Figure 5 shows the associations between the participants' previous use of apps for mental health care (Q2) and their willingness to use these apps in the future (Q3). Figure 6 presents the answers to questions Q1 to Q4 by the participants who had high or severe risk scores in Part 1 of the questionnaire.

Figure 5. Associations between the answers to Q2 and Q3.

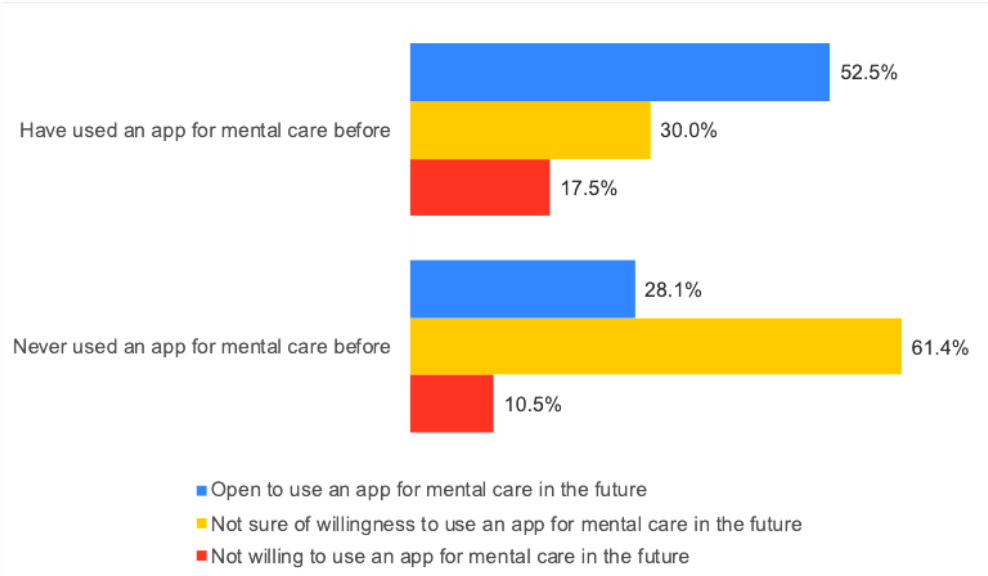
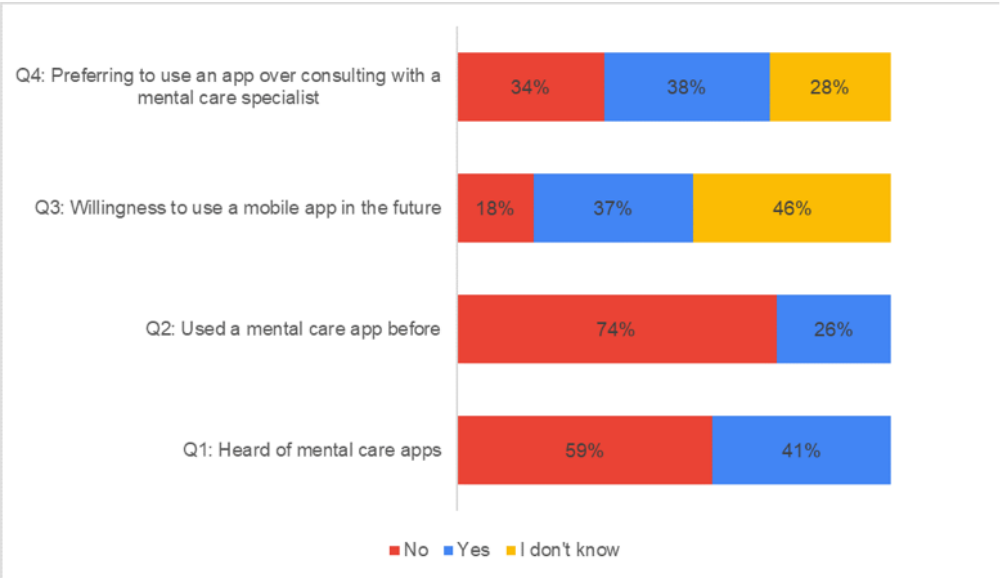


Figure 6. Answers to questions Q1 to Q4 by participants whose total GHQ-12 scores indicated high and severe mental health risk (n=68).



Discussion

Main Findings

The results of Part 1 of our questionnaire showed that more than one-third of the participants (66/154, 42.9%) were experiencing psychological health issues associated with anxiety and depression. The most frequently reported issues related to anxiety and depression were sleep problems, feelings of unhappiness, depression, and being constantly under strain. Given the period of the survey, those results can be linked to the measures taken by universities in the UAE due to the COVID-19 pandemic. The questionnaire was made available on the internet after universities were officially closed to students for two weeks and lectures were being given on the internet. Students were required to adjust to new circumstances and lifestyles and face ambiguity regarding what would happen next concerning their education and future. These findings are consistent with those of other studies conducted in other countries to analyze the psychological impact of the COVID-19 pandemic. Several studies have shown that anxiety, depression, sleep issues [33], and stress [34] are the most common psychological issues caused by the outbreak.

Many students also showed high risk of social dysfunction issues. Items 4 and 8 of the questionnaire were related to the students' capabilities of making decisions and facing problems, respectively. The nature of these issues implies that they are mainly linked to critical and problematic cases of social dysfunction; also, they may be indicators of serious pre-existing problems that may not necessarily be linked to the COVID-19 lockdown. Social dysfunction issues were mainly related to the lack of enjoyment of daily activities (item 7) and lack of a sense of playing a useful part in things (item 3). This can be linked to the pausing of many activities and the limitations on what students could do at and from home due to the safety measures taken during the outbreak.

We found weak correlations between the age of the participants and the total score of the GHQ-12 (–0.101) and between the gender of the participants and the total score of the GHQ-12 (–0.128). This indicates that university students at any age can be psychologically influenced by the lockdown due to COVID-19. The majority of participants that had high or very high psychological factor scores were female. Women have been reported to be more susceptible to anxiety, fear, and stress [35-37], which may explain this result.

The COVID-19 outbreak has caused many psychological issues among different groups of people, and with the required safety measures, it can be difficult to consult with mental health care professionals even in emergency cases. Alternative remote mental health care solutions are needed during the pandemic, just as distance learning was required as an alternative to in-school learning. Apps are convenient solutions during the current pandemic, as they provide mental health care via mobile devices, which are owned by the majority of people in the UAE, particularly the younger generation [38]. Apps for mental health care can also help overcome some pre-existing mental health care barriers, such as stigma, cost, and distance from mental health care professionals and institutions [39,40]. Apps have also been reported to be effective to alleviate anxiety, stress, and depression [15,16]. The results showed that the majority of participants (78/154, 50.6%), who are university students in the UAE, had never heard of mobile mental health care apps, and 114/154 respondents (74.0%) had never used this type of app. This was also reflected in their attitudes toward these solutions, as 82/154 participants (53.2%) were unsure about their willingness to use an app for mental health care in the future. One-third of the participants (48/154, 31.2%) were not sure if they would prefer using an app to consulting with a mental health care professional. These results were also persistent when analyzing answers of participants who had high or very high general GHQ-12 scores, as shown in Figure 6. These participants may be in need of such apps. The lack of knowledge and use of apps may be due to the fact that the younger generation mainly associates apps with games, communication, or other activities not related to health or mental health in particular [41]. A person who is not familiar with the concept of a mental health care app may not think to search for one.

There is a need for digital health literacy, particularly for mental health care, in the UAE. This can be provided using social networks such as Instagram and Twitter or via widely used websites such as YouTube, as these platforms reach a large number of people and can help spread knowledge about available mental health care apps.

Some participants (12/114, 10.5%) who had never used such apps expressed unwillingness to use one in the future; this may be due to a lack of knowledge or trust regarding these apps. Involving mental health professionals in the design of apps for mental health could give the apps more credibility and encourage university students to be more open toward using them. It is noteworthy that a group of participants who had used apps for mental health care before (7/40, 17.5%) expressed unwillingness to use them again in the future; this may be due to poor usability and/or lack of functionality of the used apps. End users should be involved in the creation of mHealth apps [42] to improve the adoption of these solutions. Integrating gamification features in these apps may also be a solution to make the apps more enjoyable and engaging and to encourage users to continue using them.

The majority of participants (109/154, 70.8%) answered "Yes" or "I don't know" when asked if they preferred using an app to consulting with a mental health care professional; these participants justified their choices as being mainly due to their lack of knowledge of mental health. Apps can be a useful means for people to educate themselves about mental issues and to obtain easy access to information. The participants also suggested that educational content could be provided in mental health care apps. It must be noted that mental health care apps are not a substitute for professional care. These apps can aid the management of certain mental issues and deliver certain treatment methods; however, seeking professional help is imperative to treat serious mental health issues.

Stigma was among the expressed reasons for preferring mental health care apps. Apps can provide anonymity and confidentiality when only users can access their content, which helps overcome stigma barriers. The aforementioned listed reasons were consistent with the students' preferred functionality and characteristics, such as web-based therapy, advice from specialists, anonymity to help overcome stigma, and affordable price to overcome cost issues. The results of this study are consistent with the results of previous studies reporting that stigma, cost, and distance from mental health care professionals are the main barriers to mental health care delivery [39,40]. Simple design and ease of use were among the characteristics suggested by the participants. They felt that the apps should be easy to use, easy to learn, and provide enjoyable user experience. The aforementioned functionality and characteristics suggested by the students could be established as requirements of mental health apps for university students in the UAE.

Although mental health care apps are a convenient means to overcome many mental health care barriers, it must be noted that for an app to be beneficial, it must ensure safety, privacy, security, and confidentiality of users' data. Available apps for mental health care differ in their quality, effectiveness, and security measures. Psychological health is a sensitive subject; therefore, the user should check the permissions required by these apps and the treatment approaches they provide before using them.

Limitations

This study may have some limitations. First, the survey was conducted during the early stages of lockdown in the UAE; the psychological state of the students may have changed since then, and they are likely to be experiencing more psychological issues. Second, the difference in the number of participants by gender may have affected the psychological investigation results; the majority of participants were female, and women have been reported to more susceptible to certain psychological issues. It should be noted that female students represent 81% of the entire UAEU student population [43], which may also have impacted the results. Third, comparing the results with a prelockdown investigation of the psychological state of university students in the UAE using the GHQ-12 would have improved the discussion of the results. However, to the best of our knowledge, no such investigation exists. Fourth, a broader number of participants may have been included if the survey had not been conducted during a pandemic. Fifth, conducting semistructured interviews could have improved the discussion of the results. However, as the study was conducted at the beginning of the lockdown due to the COVID-19 pandemic, delivering the questions through a web-based form was our best and only choice. Finally, given that more than half of the respondents were unfamiliar with mental health apps, the conclusions regarding the benefits of the use of these apps were not based on the results of the questionnaire but on results of previous studies investigating the effectiveness and advantages of mental health care apps.

Conclusions and Future Work

The COVID-19 outbreak and the applied safety measures to limit its spread have caused many psychological issues worldwide. Our psychological assessment test of university students based on the GHQ-12 showed that more than one-third of the participants were experiencing issues related to depression and anxiety as well as to social dysfunction, which confirmed our first hypothesis. In contrast, our second hypothesis was found to be unsupported, as the majority of students showed a lack of awareness of mental health care apps. The students also showed mixed attitudes and uncertain willingness to use such apps, which does not support our third hypothesis. The participants proposed characteristics and functionalities of mental health care apps that could encourage them to use such apps. We encourage developers of mental health apps to consider these suggestions, especially when targeting university students or the younger generation. We do believe that the findings of this study may assist researchers and practitioners investigating the impact of the COVID-19 outbreak on the psychological health of university students.

In future work, we intend to build on the results of this study and develop a mobile app to help university students cope with mental health issues.

Acknowledgments

We would like to thank Dr Leena Amiri for checking the suitability of the survey for university students. This study is part of the Startup project (#31T131) funded by the UAEU (2019-2021).

Authors' Contributions

All authors contributed to the creation of the manuscript. ND was involved in the study design and conception, acquisition and interpretation of data, analysis of questionnaire answers, and drafting and revision of the manuscript. SO contributed to the study design and conception, acquisition and interpretation of data, statistical support, drafting of the manuscript, and critical revision of the manuscript. MK, AH, HN, and SS were involved in the design and distribution of the questionnaire and the collection of data.

Conflicts of Interest

None declared.

References

1. Novel Coronavirus (2019-nCoV) Situation Report - 1. World Health Organization. 2020 Jan 20. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4 [accessed 2020-06-30]
2. Novel Coronavirus (2019-nCoV) Situation Report - 11. World Health Organization. 2020 Jan 31. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4 [accessed 2020-06-30]
3. Novel Coronavirus (2019-nCoV) Situation Report – 22. World Health Organization. 2020 Feb 11. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2 [accessed 2020-06-30]
4. Coronavirus disease (COVID-19) Situation Report – 104. World Health Organization. 2020 May 03. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200503-COVID-19-sitrep-104.pdf?sfvrsn=53328f46_2 [accessed 2020-06-30]
5. Coronavirus disease (COVID-19) advice for the public. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-06-30]
6. Brooks S, Webster R, Smith L, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar;395(10227):912-920. [doi: 10.1016/S0140-6736(20)30460-8]
7. Shihabuddin L. How to Manage Stress and Anxiety from Coronavirus (COVID-19). RWJBarnabas Health. 2020 Mar 13. URL: https://www.rwjbh.org/blog/2020/march/how-to-manage-stress-and-anxiety-from-coronaviru/ [accessed 2020-06-30]
8. Murray J, Sherwood H. Anxiety on rise due to coronavirus, say mental health charities. The Guardian. 2020 Mar 13. URL: https://www.theguardian.com/world/2020/mar/13/anxiety-on-rise-due-to-coronavirus-say-mental-health-charities [accessed 2020-06-30]
9. Zulkefly SN, Baharudin R. Using the 12-item General Health Questionnaire (GHQ-12) to Assess the Psychological Health of Malaysian College Students. GJHS 2010 Mar 17;2(1):73. [doi: 10.5539/gjhs.v2n1p73]
10. Baksheev GN, Robinson J, Cosgrave EM, Baker K, Yung AR. Validity of the 12-item General Health Questionnaire (GHQ-12) in detecting depressive and anxiety disorders among high school students. Psychiatry Res 2011 May 15;187(1-2):291-296. [doi: 10.1016/j.psychres.2010.10.010] [Medline: 21067813]
11. Yaghubi H. Validity and factor structure of the General Health Questionnaire (GHQ-12) in university students. IJBS 2012;6(2):153-160 [FREE Full text]
12. Mari JJ, Williams P. A comparison of the validity of two psychiatric screening questionnaires (GHQ-12 and SRQ-20) in Brazil, using Relative Operating Characteristic (ROC) analysis. Psychol Med 1985 Aug 09;15(3):651-659. [doi: 10.1017/s0033291700031500] [Medline: 4048323]
13. Picardi A, Abeni D, Pasquini P. Assessing psychological distress in patients with skin diseases: reliability, validity and factor structure of the GHQ-12. J Eur Acad Dermatol Venereol 2001 Sep;15(5):410-417. [doi: 10.1046/j.1468-3083.2001.00336.x] [Medline: 11763380]
14. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan 01;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]
15. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of Mobile Apps for Depression and Anxiety: Cross-Sectional Survey Study on Patient Interest and Barriers to Engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
16. Coulon SM, Monroe CM, West DS. A Systematic, Multi-domain Review of Mobile Smartphone Apps for Evidence-Based Stress Management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]

17. Gaupp-Berghausen M, Raser E, Anaya-Boig E, Avila-Palencia I, de Nazelle A, Dons E, et al. Evaluation of Different Recruitment Methods: Longitudinal, Web-Based, Pan-European Physical Activity Through Sustainable Transport Approaches (PASTA) Project. J Med Internet Res 2019 Mar 09;21(5):e11492 [FREE Full text] [doi: 10.2196/11492] [Medline: 31066715]
18. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
19. Goldberg DP. The detection of psychiatric illness by questionnaire: A technique for the identification and assessment of non-psychotic psychiatric illness. Oxford, UK: Oxford University Press; 1972.
20. Buck N, Gershuny J, Rose D, Scott J, editors. Changing Households: The British Household Panel Survey 1990 -1992. Essex, UK: ESRC Research Centre on Micro-Social Change, University of Essex; 1994.
21. Worsley A, Gribbin CC. A factor analytic study on the twelve item general health questionnaire. Aust N Z J Psychiatry 1977 Dec 26;11(4):260-272. [doi: 10.3109/00048677709159577] [Medline: 272886]
22. Andrich D, van Schoubroeck L. The General Health Questionnaire: a psychometric analysis using latent trait theory. Psychol Med 1989 May 09;19(2):469-485. [doi: 10.1017/s0033291700012502] [Medline: 2762448]
23. Politi PL, Piccinelli M, Wilkinson G. Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. Acta Psychiatr Scand 1994 Dec;90(6):432-437. [doi: 10.1111/j.1600-0447.1994.tb01620.x] [Medline: 7892776]
24. Martin AJ. Assessing the multidimensionality of the 12-item General Health Questionnaire. Psychol Rep 1999 Jun 01;84(3 Pt 1):927-935. [doi: 10.2466/pr0.1999.84.3.927] [Medline: 10408215]
25. Mäkikangas A, Feldt T, Kinnunen U, Tolvanen A, Kinnunen M, Pulkkinen L. The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: evidence from two community-based samples. Psychol Assess 2006 Dec;18(4):444-451. [doi: 10.1037/1040-3590.18.4.444] [Medline: 17154766]
26. Cheung YB. A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. Int J Geriatr Psychiatry 2002 Aug;17(8):739-744. [doi: 10.1002/gps.693] [Medline: 12211124]
27. Gao F, Luo N, Thumboo J, Fones C, Li S, Cheung Y. Does the 12-item General Health Questionnaire contain multiple factors and do we need them? Health Qual Life Outcomes 2004 Nov 11;2:63 [FREE Full text] [doi: 10.1186/1477-7525-2-63] [Medline: 15538951]
28. General Health Questionnaire. GL Assessment. 2020. URL: https://www.gl-assessment.co.uk/products/general-health-questionnaire-ghq/ [accessed 2020-06-30]
29. Sreejith G, Menon V. Mobile Phones as a Medium of Mental Health Care Service Delivery: Perspectives and Barriers among Patients with Severe Mental Illness. Indian J Psychol Med 2019 Oct 01;41(5):428-433. [doi: 10.4103/ijpsym.ijpsym_333_18]
30. Sukmawati I, Ardi Z, Ifdil I, Zikra Z. Development and Validation of Acceptability of Mental-Health Mobile App Survey (AMMS) for Android-based Online Counseling Service Assessment. J Phys Conf Ser 2019 Dec 16;1339:012124. [doi: 10.1088/1742-6596/1339/1/012124]
31. Miller KE, Kuhn E, Yu J, Owen JE, Jaworski BK, Taylor K, et al. Use and perceptions of mobile apps for patients among VA primary care mental and behavioral health providers. Prof Psychol Res Pr 2019 Jun;50(3):204-209. [doi: 10.1037/pro0000229]
32. Atallah N, Khalifa M, El Metwally A, Househ M. The prevalence and usage of mobile health applications among mental health patients in Saudi Arabia. Comput Methods Programs Biomed 2018 Mar;156:163-168. [doi: 10.1016/j.cmpb.2017.12.002] [Medline: 29428068]
33. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
34. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
35. Bhui K, Fletcher A. Common mood and anxiety states: gender differences in the protective effect of physical activity. Soc Psychiatry Psychiatr Epidemiol 2000 Jan 16;35(1):28-35. [doi: 10.1007/s001270050005] [Medline: 10741533]
36. Keogh E, Hamid R, Hamid S, Ellery D. Investigating the effect of anxiety sensitivity, gender and negative interpretative bias on the perception of chest pain. Pain 2004 Sep;111(1-2):209-217. [doi: 10.1016/j.pain.2004.06.017] [Medline: 15327825]
37. Lewinsohn PM, Gotlib IH, Lewinsohn M, Seeley JR, Allen NB. Gender differences in anxiety disorders and anxiety symptoms in adolescents. J Abnorm Psychol 1998;107(1):109-117. [doi: 10.1037/0021-843x.107.1.109]
38. Smartphone penetration rate in the MENA 2018, by selected country. Statista. 2019 May 14. URL: https://www.statista.com/statistics/1005783/mena-smartphone-penetration-rate-by-selected-country/ [accessed 2020-06-30]
39. Lingley-Pottie P, McGrath PJ, Andreou P. Barriers to Mental Health Care. Adv Nurs Sci 2013;36(1):51-61. [doi: 10.1097/ans.0b013e31828077eb]
40. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2013 Aug 09;44(6):1303-1317. [doi: 10.1017/s0033291713001943]

41. Middle East: daily life activities via smartphone 2016 Published by Statista Research Department, Jun 10, 2016 This statistic shows the frequency of life activities internet-going smartphone users performed daily in the Middle East, in 2016, by select country. During the survey period it was found that 73 percent of internet-going smartphone users in the UAE took photos via smartphone daily. Daily life activities performed on smartphones in the Middle East, as of 2016, by select country. Statista. 2016 Jun 10. URL: https://www.statista.com/statistics/731880/daily-life-activities-middle-east/ [accessed 2020-06-30]
42. Ouhbi S, Karampela M, Isomursu M. Integrating Users Logic Into Requirements Engineering for Connected Healthcare co-Design. In: Proceedings of the 14th International Conference on Evaluation of Novel Approaches to Software Engineering - Volume 1: ENASE. 2019 Presented at: 14th International Conference on Evaluation of Novel Approaches to Software Engineering; May 4-5, 2019; Heraklion, Crete, Greece p. 480-485. [doi: 10.5220/0007754504800485]
43. United Arab Emirates University. URL: https://www.uaeu.ac.ae/en/about/facts_and_figures.shtml [accessed 2020-06-30]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
GHQ-12: 12-item General Health Questionnaire
GHQ-28: 28-item General Health Questionnaire
mHealth: mobile health
UAE: United Arab Emirates
UAEU: United Arab Emirates University
WHO: World Health Organization

Edited by J Torous, G Eysenbach; submitted 05.05.20; peer-reviewed by N Lau, V Strotbaum, K Reynolds; comments to author 26.06.20; revised version received 09.07.20; accepted 15.09.20; published 20.10.20.

Please cite as:
Drissi N, Alhmoudi A, Al Nuaimi H, Alkhyeli M, Alsalami S, Ouhbi S
Investigating the Impact of COVID-19 Lockdown on the Psychological Health of University Students and Their Attitudes Toward Mobile Mental Health Solutions: Two-Part Questionnaire Study
JMIR Form Res 2020;4(10):e19876
URL: http://formative.jmir.org/2020/10/e19876/
doi:10.2196/19876
PMID:32969340

Original Paper

Mental Health During the COVID-19 Pandemic in the United States: Online Survey

Jennifer S Jewell[1], BS, MSPH; Charlotte V Farewell[1], PhD; Courtney Welton-Mitchell[1], PhD; Angela Lee-Winn[1], PhD; Jessica Walls[1], BA; Jenn A Leiferman[1], PhD

Colorado School of Public Health, Aurora, CO, United States

Corresponding Author:
Jennifer S Jewell, BS, MSPH
Colorado School of Public Health
Building 500
13001 E 17th Place
Aurora, CO
United States
Phone: 1 303 519 6620
Email: jennifer.jewell@cuanschutz.edu

Abstract

Background: The COVID-19 pandemic has had numerous worldwide effects. In the United States, there have been 8.3 million cases and nearly 222,000 deaths as of October 21, 2020. Based on previous studies of mental health during outbreaks, the mental health of the population will be negatively affected in the aftermath of this pandemic. The long-term nature of this pandemic may lead to unforeseen mental health outcomes and/or unexpected relationships between demographic factors and mental health outcomes.

Objective: This research focused on assessing the mental health status of adults in the United States during the early weeks of an unfolding pandemic.

Methods: Data was collected from English-speaking adults from early April to early June 2020 using an online survey. The final convenience sample included 1083 US residents. The 71-item survey consisted of demographic questions, mental health and well-being measures, a coping mechanisms checklist, and questions about COVID-19–specific concerns. Hierarchical multivariable logistic regression was used to explore associations among demographic variables and mental health outcomes. Hierarchical linear regression was conducted to examine associations among demographic variables, COVID-19–specific concerns, and mental health and well-being outcomes.

Results: Approximately 50% (536/1076) of the US sample was aged ≥45 years. Most of the sample was White (1013/1054, 96%), non-Hispanic (985/1058, 93%), and female (884/1073, 82%). Participants reported high rates of depression (295/1034, 29%), anxiety (342/1007, 34%), and stress (773/1058, 73%). Older individuals were less likely to report depressive symptomology (OR 0.78, $P<.001$) and anxiety symptomology (OR 0.72, $P<.001$); in addition, they had lower stress scores (–0.15 points, SE 0.01, $P<.001$) and increased well-being scores (1.86 points, SE 0.22, $P<.001$). Individuals who were no longer working due to COVID-19 were 2.25 times more likely to report symptoms of depression ($P=.02$), had a 0.51-point increase in stress (SE 0.17, $P=.02$), and a 3.9-point decrease in well-being scores (SE 1.49, $P=.009$) compared to individuals who were working remotely before and after COVID-19. Individuals who had partial or no insurance coverage were 2-3 times more likely to report depressive symptomology compared to individuals with full coverage ($P=.02$ and $P=.01$, respectively). Individuals who were on Medicare/Medicaid and individuals with no coverage were 1.97 and 4.48 times more likely to report moderate or severe anxiety, respectively ($P=.03$ and $P=.01$, respectively). Financial and food access concerns were significantly and positively related to depression, anxiety, and stress (all $P<.05$), and significantly negatively related to well-being (both $P<.001$). Economy, illness, and death concerns were significantly positively related to overall stress scores (all $P<.05$).

Conclusions: Our findings suggest that many US residents are experiencing high stress, depressive, and anxiety symptomatology, especially those who are underinsured, uninsured, or unemployed. Longitudinal investigation of these variables is recommended. Health practitioners may provide opportunities to allay concerns or offer coping techniques to individuals in need of mental health care. These messages should be shared in person and through practice websites and social media.

(JMIR Form Res 2020;4(10):e22043) doi:10.2196/22043

KEYWORDS

COVID-19; mental health; pandemic; depression; anxiety; well-being; stress

Introduction

The COVID-19 pandemic has produced over 41 million confirmed cases and over 1.1 million confirmed deaths worldwide as of October 21, 2020 [1]. Of these, nearly 8.3 million cases are in the United States, with nearly 222,000 deaths [1]. In addition to health impacts, many have raised the alarm about the potential for a widespread global mental health crisis as a result of the pandemic [2-5]. Specific groups may be at increased risk for adverse mental health outcomes, such as frontline health care workers [6] and those that have experienced illness or death of family, friends, or coworkers. Many more are likely to experience distress as a result of economic hardship, disruption to social networks, and work- and school-related changes due to the protracted crisis.

Elevated rates of depression and anxiety have been documented following stressors such as disease outbreaks, including the 2014-2016 Ebola crisis in West Africa, among caretakers, survivors, their immediate contacts, and others [7,8]. In addition, epidemics such as SARS and HIV have been associated with depression and other mental health concerns among various groups [9-14]. The current pandemic is likely to be associated with similar mental health outcomes, as a result of potential exposure to stressors including loss of loved ones, economic hardship, social isolation, and childcare responsibilities following school and day care closures.

Countless businesses across the United States closed in an attempt to protect workers, limit transmission of the coronavirus, and allow health care systems to keep pace with the needs of those requiring hospital care. With the exception of essential services, much of the economy has come to a virtual standstill, resulting in unprecedented rates of unemployment [15]. Financial struggles, including job loss and food insecurity, are known risk factors for mental illness, particularly anxiety, depression, and suicide [16,17].

In most US states, nonessential workers have been required to stay at home for several weeks. Many states have had stay-at-home orders in place for longer periods of time. Although there is an easing of movement restrictions in some areas within the United States, many people are still concerned about the potential safety risks of resuming prepandemic levels and types of activities. As a result, so-called “social distancing” continues for many in the United States. Physical distancing requirements (eg, social distancing) have the potential to limit physical and social contact, disrupt prepandemic social networks, and undermine the potential for social support at a time when it may be needed most. This may result in an increase in loneliness and social isolation. Across numerous studies, social isolation has been associated with increased morbidity and mortality, with an increase in coronary heart disease, stroke, and poor mental health outcomes such as depression and anxiety [18-22].

The increase in financial and familial struggles for some families may have exacerbated the negative effects of strict social distancing measures and overall trauma. Although studies examining the mental health impacts of COVID-19 are limited, findings from a few recent studies indicate that many in the United States are experiencing significant and worsening mental health difficulties during the pandemic [23]. A review of the emerging literature regarding the effects of the pandemic suggests that symptoms of anxiety and depression are common [24]. In one study [25], which used a representative sample and compared recent mental health concerns to those in 2018, large increases in mental health distress were noted. Younger people, those with children in the household, married individuals, and Asians appeared to be faring worse than others [25]. Authors suggested these findings may reflect economic hardship, but more research is needed to understand factors contributing to greater difficulties in some groups than others.

The current study examines demographic differences in mental health and well-being outcomes and specific sources of concern that impact these outcomes among a US sample of 1083 adults surveyed between April 7 and June 1, 2020, immediately following business closures and movement restrictions. This study may bring to light additional factors related to mental health during the pandemic and fill gaps in the current literature. Specifically, several COVID-19–specific concern-related items that have not been previously assessed were included in the current analyses. These findings have the potential to inform current intervention efforts as well as new initiatives, with the potential to mitigate suffering and bolster resilience during the ongoing pandemic.

Methods

Procedures

The Mental Health and Wellbeing Survey during COVID-19 Pandemic received ethical approval from the Colorado Multiple Institutional Review Board (COMIRB Protocol #20-0676). Survey data was collected between April 7 and June 1, 2020. A snowball sampling technique was used. This survey was advertised on Facebook and Instagram via paid targeted advertising. In addition, it was sent out via listservs and other media including Centers for Disease Control and Prevention (CDC) Prevention Research Centers, American Public Health Association Mental Health Section, Colorado Public Radio, University of Colorado research announcements, and the University of South Florida. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Colorado [26]. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing the following features: (1) an intuitive interface for validated data entry, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for importing data from external sources. Participants consented digitally before beginning the survey. Additionally, participants in the initial survey were given the opportunity to opt in to future

surveys to collect longitudinal data. A participation incentive in the form of a drawing for one of two $50 gift cards was offered.

Participants

Adults aged ≥18 years were eligible to take the English-language survey, regardless of country of residence. There were no exclusion criteria beyond ability to provide consent. Although data was collected from an international sample initially, most of the participants were residing in the United States. As a result, only data from the US subsample is included in the present analyses. The final US sample consisted of 1083 individuals.

Measures

The 71-item survey consisted of demographic questions, mental health and well-being measures, coping mechanisms, and questions gauging COVID-19–specific concerns. Demographic questions included age, race/ethnicity, gender, work status, household size, and insurance coverage. The survey also included four mental health and well-being scales measuring well-being, depression, anxiety, and stress. The Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS) was used as a continuous measure of well-being. It has high internal consistency and convergent validity with other measures of life satisfaction and physical and mental health (α=.93 in this sample). The SWEMWBS has a range of 7-35, with higher scores indicating higher well-being [27]. The Patient Health Questionnaire-2 (PHQ-2) was used as a brief measure of depression (α=.81 in this sample). The PHQ-2 has a sensitivity of 83% and a specificity of 92% for major depression. The PHQ-2 has a range of 0-6 and was dichotomized for analyses using a cutoff score of ≥3 [28,29]. Generalized Anxiety Disorder (GAD) was assessed using the GAD-7, which has a sensitivity of 89% and a specificity of 82% (α=.92 in this sample). The GAD-7 has a range of 0-21, and moderate or severe anxiety was based on a cut-off of ≥10 [30]. Lastly, stress was assessed using a validated 1-item continuous measure with 5 response options ranging from “not at all” to “very much” stress “these days” (Elo stress-symptoms item). This stress item has demonstrated construct, content, and criterion validity for group-level analysis [31].

The survey included a coping checklist, comprised of 12 behavioral items with an additional “other” option, to ascertain which types of coping were most common (eg, exercise, engaging with media, engaging remotely with family/friends). The survey items examining COVID-19–specific concerns included questions about personal financial impact, food security, economic impact, and risk of serious illness or death (in participants or others known to participants) related to COVID-19. Questions were phrased in the following manner: “How concerned are you about... [the financial impact current events may have on your family]?”

Analyses

Data were exported from REDCap into SPSS (Version 25; IBM Corp) for analyses. Data cleaning included testing of assumptions, exploration of outliers, and missingness for all key variables. As all key variables had less than 10% missing data and data were missing completely at random (χ^2_9=12.86, P=.17), listwise deletion was used in all analyses. Univariate and bivariate analyses were conducted. Two proportion z tests were also used to calculate differences between responses (%) to the PHQ-2 and GAD-7 and national prevalence data. An independent sample t test was run to compare the sample average for the Warwick wellbeing score with a nationally representative sample.

Two hierarchical multivariable logistic regression models were run (logistic regression models 1 and 2) to explore associations among demographic variables, depression (not depressed versus depressed), and anxiety (no or mild anxiety versus moderate or severe anxiety) outcomes. Hierarchical regression was used to investigate if specific sources of concern (eg, financial concern, illness-related concern) were related to the outcome measures after controlling for demographic characteristics of the analytical sample. For categorical variables, well-established cutoffs based on representative US samples were used. All demographic variables were added simultaneously to each model, after which 5 unique sources of concern were entered into models (logistic regression models 3 and 4) to see which sources of concern predicted depression and anxiety outcomes after controlling for demographics. R^2 values, odds ratios, and P values for logistic regression models are presented.

Next, two hierarchical linear regression models were run (linear regression models 1 and 2) to explore associations between demographic variables and stress and well-being outcomes. In total, 5 unique sources of concern were entered into models (linear regression models 3 and 4) to see which sources of concern predicted stress and well-being outcomes after controlling for demographics. Unstandardized coefficients, P values, and adjusted R^2 values are reported for all linear regression models. Alpha (α) was set at .05.

Results

Table 1 depicts demographics and the prevalence of mental health and well-being indicators of the final analytical sample, which included 1083 individuals. In total, 45 states within the United States were represented within the sample. Overall, 56% of the participants resided in Colorado. The remaining states comprised 0%-4% of the sample. Approximately 50% (536/1076) of the analytical sample were aged ≥45 years. Most of the sample was White (1013/1054,96%). Hispanic individuals made up 7% (73/1058) of the sample and 82% (884/1073) of participants were female. The average household size of the sample was 2.6 individuals. The self-reported depression rate in the sample population was 29% (295/1034) compared to a national average of 7% (z=27.8, P<.01) [32]. Approximately one-third of the sample reported moderate or severe anxiety (342/1007, 34%) compared to a national average of 20% of US adults prior to the pandemic (z=11.4, P<.01) [33]. Three-quarters (773/1058, 73%) of the sample reported experiencing stress “to some extent” or greater (“rather much” or “very much”). The average well-being score of the sample was 45.1 (SD 10.0), which compares to a national average of 51 (t[4349]=17.02, P<.01) [34]. From the 12 items provided (including the “other” specify option), the most prevalent coping behaviors reported by the sample included use of television (661/1083, 61%),

texting with family and friends (661/1083, 61%), social media (617/1083, 57%), and exercise (617/1083, 57%).

Logistic regression models for the depression and anxiety outcomes are presented in Table 2. Age was related to mental health outcomes; older individuals were less likely to report depressive and anxiety symptomology compared to younger individuals (OR 0.78, 95% CI 0.70-0.87, *P*<.001 and OR 0.72, 95% CI 0.65-0.80, *P*<.001, respectively). Additionally, individuals who were no longer working due to COVID-19 were 2.25 times more likely to report symptoms of depression compared to individuals who were working remotely before and after COVID ("no change" group; 95% CI 1.15-4.43, *P*=.02). Finally, insurance status was significantly associated with both depression and anxiety outcomes. Individuals who had partial coverage and individuals with no coverage were 2.67 and 3.22 times more likely to report depressive symptomology compared to individuals with full coverage, respectively (95% CI 1.91-6.00, *P*=.02 and 95% CI 1.33-7.80, *P*=.01, respectively). Individuals who were on Medicare/Medicaid and individuals with no coverage were 1.97 and 4.48 times more likely to report moderate or severe anxiety compared to individuals with full coverage, respectively (95% CI 1.09-3.57, *P*=.03 and 95% CI 1.73-11.60, *P*=.01, respectively).

Linear regression models for the well-being and stress outcomes are presented in Table 3. An increase in age decade was associated with a 0.15-point decrease in stress score (SE 0.01, *P*<.001) and a 1.86-point increase in well-being score (SE 0.22, *P*<.001). On average, individuals who did not have insurance reported a 0.72-point higher stress score (SE 0.29, *P*=.002) and a 9.59-point lower well-being score (SE 2.09, *P*<.001). No longer working due to COVID-19 was associated with a 0.51-point increase in stress score and 3.90-point decrease in well-being score compared to individuals who were working remotely before and after COVID ("no change" group; SE 0.17, *P*=.02; SE 1.49, *P*=.009). Males also reported significantly lower stress scores compared to females (B=0.42, SE 0.10, *P*<.001).

Financial concerns and food access concerns were significantly and positively related to depression, anxiety, and stress (all *P*<.05) and significantly negatively related to well-being (both *P*<.001). Economy-, illness-, and death-related concerns were significantly and positively related to overall stress score after controlling for all demographic variables (all *P*<.05).

Additional analyses were considered, including investigating the effects of race/ethnicity and parenthood status. The cell sizes for these variables were too small to conduct analyses.

Table 1. Sample characteristics and mental health and well-being (N=1083).

Variables	Values
Demographics	
Age (years), n (%)	
18-25	46 (4.3)
25-44	494 (45.9)
45-59	313 (29.1)
≥60	223 (20.7)
Race, n (%)	
White	1013 (96.1)
Black or African American	4 (0.4)
Asian	13 (1.2)
Native Hawaiian or Pacific Islander	24 (2.3)
Ethnicity, n (%)	
Hispanic	73 (6.9)
Non-Hispanic	985 (93.1)
Gender, n (%)	
Male	189 (17.6)
Female	884 (82.4)
Health care insurance, n (%)	
Full coverage	893 (82.6)
Partial coverage	38 (3.5)
Medicaid/Medicare	120 (11.1)
No insurance	30 (2.8)
Work status, n (%)	
Remote before and after COVID-19	104 (9.9)
Unemployed prior to COVID-19	167 (15.9)
Work outside home	120 (11.4)
No longer working due to COVID-19	107 (10.2)
Working remotely due to COVID-19	552 (11.4)
Household size, mean (SD)	2.6 (1.4)
Mental health and well-being variables	
Warwick Wellbeing Scale, mean (SD)	45.1 (10.0)
Anxiety (GAD-7[a]), n (%)	
No/mild (<10)	665 (66.0)
Moderate or severe (≥10)	342 (34.0)
Depression (PHQ-2[b]), n (%)	
Not depressed (<3)	739 (71.5)
Depressed (≥3)	295 (28.5)
Stress, n (%)	
Not at all	67 (6.3)
Only a little	218 (20.6)
To some extent	363 (34.3)

Variables	Values
Rather much	240 (22.7)
Very much	170 (16.1)

[a]GAD-7: Generalized Anxiety Disorder 7-item scale.

[b]PHQ-2: Patient Health Questionnaire 2-item scale.

Table 2. Logistic regression models showing associations between depression (Models 1 and 3), anxiety (Models 2 and 4), demographic variables, and sources of concern (N=1083).

Predictor variables	Model 1 (PHQ-2[a])		Model 2 (GAD-7[b])		Model 3 (PHQ-2)		Model 4 (GAD-7)	
	Odds ratio	*P* value	Odds ratio	*P* value	Odds ratio	*P* value	Odds ratio	*P* value
Age	0.78	<.001[c]	0.72	<.001[c]	0.76	<.001[c]	0.68	<.001[c]
Race								
White	Ref[d]	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Black	2.13	.45	2.00	.49	0.75	.82	0.54	.64
Native American or American Indian	1.91	.31	2.11	.24	1.98	.32	2.35	.21
Asian	0.50	.24	0.41	.14	0.49	.25	0.34	.10
Ethnicity								
Hispanic	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Non-Hispanic	1.34	.39	1.02	.94	1.43	.31	1.01	.99
Work status								
Working remotely before and after COVID-19	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Unemployed prior to COVID-19	1.51	.22	0.92	.79	1.33	.42	0.69	.26
Work outside home	1.00	.99	0.97	.93	1.00	.99	0.91	.78
No longer working due to COVID-19	2.25	.02[e]	1.92	.06	1.80	.11	1.32	.45
Working remotely due to COVID-19	0.75	.31	0.77	.36	0.72	.28	0.70	.22
Insurance								
Full coverage	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Partial coverage	2.67	.02[e]	1.20	.67	2.35	.04[e]	1.05	.91
Medicare/Medicaid	1.22	.49	1.97	.03[e]	1.19	.58	2.18	.02[e]
None	3.22	.01[c]	4.48	<.001[c]	1.84	.20	3.09	.03[e]
Gender								
Female	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Male	1.19	.41	0.75	.19	1.33	.22	0.91	.68
Household size	0.91	.12	0.97	.65	0.91	.13	0.99	.84
Financial concern	N/A[f]	N/A	N/A	N/A	1.49	<.001[c]	1.32	.01[c]
Food access concern	N/A	N/A	N/A	N/A	1.29	.01[c]	1.39	<.001[c]
Economy-related concern	N/A	N/A	N/A	N/A	1.27	.08	1.29	.07
Illness-related concern	N/A	N/A	N/A	N/A	0.97	.84	1.33	.12
Death-related concern	N/A	N/A	N/A	N/A	1.21	.21	1.33	.07
Adjusted R^2	.07	N/A	.09	N/A	.13	N/A	.18	N/A

[a]PHQ-2: Patient Health Questionnaire 2-item scale.

[b]GAD-7: Generalized Anxiety Disorder 7-item scale.

[c]*P*<.05.

[d]Ref: reference.

[e]*P*<.01.

[f]N/A: not applicable.

Table 3. Linear regression models showing associations between stress (Models 1 and 3), well-being (Models 2 and 4), demographic variables, and sources of concern (N=1083).

Predictor variables	Model 1 (stress)	Model 2 (well-being)	Model 3 (stress)	Model 4 (well-being)
	B (SE)	B (SE)	B (SE)	B (SE)
Age	–0.15 (.02)[a]	1.86 (0.22)[a]	–0.15 (0.02)[a]	1.91 (0.21)[a]
Race				
White	Reference	Reference	Reference	Reference
Black	0.41 (0.54)	2.62 (4.71)	0.12 (0.55)	3.57 (5.23)
Native American or American Indian	0.36 (0.33)	2.18 (2.97)	0.33 (0.29)	2.30 (2.76)
Asian	–0.32 (0.25)	1.94 (2.24)	–0.32 (0.22)	2.15 (2.16)
Ethnicity				
Hispanic	Reference	Reference	Reference	Reference
Non-Hispanic	–0.08 (0.54)	2.62 (1.48)	–0.03 (0.15)	2.06 (1.44)
Work status				
Working remotely before and after COVID-19	Reference	Reference	Reference	Reference
Unemployed prior to COVID-19	–0.03 (0.16)	–2.81 (1.41)	–0.12 (0.14)	–2.09 (1.36)
Work outside home	0.04 (0.16)	0.03 (1.47)	0.02 (0.15)	0.12 (1.43)
No longer working due to COVID-19	0.51 (0.17)[a]	–3.90 (1.49)[a]	0.26 (0.15)	–2.58 (1.46)
Working remotely due to COVID-19	0.07 (0.13)	–0.33 (1.19)	0.05 (0.12)	–0.20 (1.16)
Insurance				
Full coverage	Reference	Reference	Reference	Reference
Partial coverage	0.50 (0.21)[b]	–2.20 (1.19)	0.37 (0.19)[b]	–1.57 (1.87)
Medicare/Medicaid	0.24 (0.14)	–1.77 (1.26)	0.21 (0.12)	–1.70 (1.22)
None	0.72 (0.29)[b]	–9.59 (2.09)[a]	0.28 (0.21)	–6.42 (2.06)[b]
Gender				
Female	Reference	Reference	Reference	Reference
Male	–0.42 (0.10)[a]	1.61 (0.91)	–0.28 (0.09)[b]	1.17 (0.89)
Household size	–0.05 (0.03)	0.44 (0.25)	–0.04 (0.03)	0.50 (0.25)[b]
Financial concern	N/A[c]	N/A	0.28 (0.04)[a]	–1.41 (0.40)[a]
Food access concern	N/A	N/A	0.18 (0.04)[a]	–1.77 (0.42)[a]
Economy-related concern	N/A	N/A	0.15 (0.05)[b]	–0.79 (0.49)
Illness-related concern	N/A	N/A	0.15 (0.07)[b]	0.94 (0.69)
Death-related concern	N/A	N/A	0.13 (0.06)[b]	–1.37 (0.61)[b]
Intercept	4.21 (0.20)	33.94 (1.81)	1.79 (0.26)	43.63 (2.56)
Adjusted R^2	.10	.12	.29	.19

[a]P<.05.

[b]P<.01.

[c]N/A: not applicable.

Discussion

The imposed social distancing experienced by many throughout the United States undoubtedly contributed to numerous short- and long-term negative effects within the population. This survey aimed to identify the impact of the COVID-19 pandemic and imposed social distancing on mental health among US residents within a small window of time during which many businesses were closed and many individuals were out of work. Based on the findings associated with this convenience sample,

when compared to prepandemic representative population-level data in the United States, it appears that mental health declined overall during the late spring of 2020. Prevalence rates of both depressive symptoms and anxiety symptoms were notably higher than national prepandemic averages. In addition, mental well-being significantly decreased, and stress levels were elevated in this sample. These findings support early evidence that the effects of the pandemic on mental health are significant [23].

The findings from the regression analyses suggest that age may be an important factor in considering mental health impacts of the pandemic. As age increased, anxiety symptoms, depression symptoms, and stress decreased, and well-being increased. This effect may be explained by stress on younger individuals due to inconsistent income or parenting-related obligations; however, these relationships could not be analyzed due to small cell sizes. Based on a review of the limited literature specifically related to the COVID-19 pandemic, Rajkumar [24] found that older adults were at greater risk for mental health concerns [35]. No other studies we reviewed found a relationship with age. Further research should be conducted to determine mental health risks relative to age and associated factors during the COVID-19 pandemic.

Findings from this study suggest loss of work due to pandemic-related closures greatly increased the odds of depression symptoms when compared to individuals who did not experience a change in their employment (were working remotely both before and after closures began). Loss of employment was also related to increased stress levels and decreased mental well-being. This could indicate a segment of the population that may require additional support to overcome mental health challenges during the pandemic. Economic crises have been tied to poor mental health outcomes in numerous studies [16,17]. Employment, in contrast to unemployment, has been linked to decreased mental illness, including depression and anxiety, and increased mental well-being [36]. Job instability, including moving from a permanent position to a temporary position, has been linked to increased mental illness [37]. Public health officials should make targeted efforts to reach out to the segment of the population that completely lost the ability to work during social distancing regulations. These individuals may need aid that extends beyond financial support.

Partial and no insurance coverage was associated with increased odds of depression symptoms when compared to fully insured individuals. This finding supports previous evidence that increased health care coverage reduces the prevalence of undiagnosed and untreated depression [38]. Individuals with limited health coverage also had higher stress scores and lower well-being scores. A similar effect was seen with moderate to severe anxiety. This finding was particularly pronounced in the uninsured population. The effects of partial or no insurance coverage on mental health may be exacerbated by the circumstances of the pandemic. Those with no insurance demonstrated extremely high odds of anxiety symptoms. This is likely related to concern about what would happen to them if they contracted COVID-19. Practitioners working with uninsured and partially insured individuals should take note of potentially decreased mental health in this population. Although these practitioners may not have the ability to affect their patients' insurance status or concerns about the potential financial burden of contracting COVID-19, they do have the opportunity to encourage low- or no-cost coping methods that may decrease depressive and anxiety symptomatology.

Several other factors demonstrated relationships with mental health. Males reported significantly lower stress levels than females. This is consistent with findings on gender and stress [39]. This difference in stress levels may be due to gender differences in coping with stressful situations and differences in hormonal responses to stressful events [40]. Increased family financial concern and family food access concern were positively related with depression symptoms, anxiety symptoms, and stress, and negatively related to well-being. In addition, concern about the economy, illness-related concern, and death-related concern were positively related to stress scores. The financial concern and food security findings are consistent with previous work investigating this relationship [41,42]. Each of the relationships between the concern items and mental health variables is consistent with expected outcomes from the COVID-19 pandemic [43]. Practitioners may wish to ask their patients about specific concerns that they may be experiencing during this time. Using a sliding scale for medical fees and having referrals and information about different types of aid available (eg, food banks and local, state, and federal funds) may reduce the mental burden on some individuals. Practitioners are also in the best position to convey accurate information about COVID-19 risk status and effective protective measures. Information of this type can be conveyed in person or online through practice websites and social media. This reliable information may counteract the concern of illness and death and reduce poor mental health outcomes.

There are noteworthy limitations to this study. The convenience sample was primarily insured, non-Hispanic, White, and female, which may have led to results that are not generalizable to the broader population of US adults. Minority populations tend to experience the effects of trauma to a greater degree than others. Given the results seen in this study in a non-Hispanic White population that is primarily insured, it is reasonable to assume that minority populations may be impacted to an even greater degree than what was demonstrated in this study. Particular care should be taken to measure and address these concerns in future studies.

In addition, due to the small number of African Americans in this sample, we were not able to explore the relationship between race and mental health, a limitation that should be prioritized for exploration in follow-up research. In addition, the sample did not include a representative percentage of young people or individuals with children. Given the age effects in this study, further investigation is encouraged to determine the effect of age on mental health outcomes during the pandemic. The results of this study are based on a comparison with prepandemic norms, which may not be representative of the morbidity of these mental health conditions in peripandemic or postpandemic times. Functional impairment was not measured. Therefore, assumptions about the impact of negative mental health symptomatology in the peripandemic period cannot be made. Furthermore, the survey was conducted online, which likely

inadvertently excluded individuals that do not have access to or are uncomfortable with the internet.

The strengths of this study include the large sample, which consisted of respondents from 45 of 50 states in the United States. This survey was also developed and launched early in the pandemic's course through the United States. Therefore, it likely captured early mental health responses that later surveys may not have captured. These responses included both mental health struggles and positive mental health indicators. This study was designed with a follow-up in mind. Respondents to this survey were asked if they would be willing to participate in a follow-up survey at a later date. This will allow for longitudinal data collection at multiple time points as social distancing restrictions change throughout the United States.

Our findings suggest that many US citizens, particularly non-Hispanic, White, insured individuals, are experiencing high stress, depressive, and anxiety symptomatology. Practitioners, including health care workers and mental health specialists, can be a resource for those struggling with mental health concerns during the pandemic. These messages should not only be made in person, but also through practice websites and social media accounts. The overwhelming amount of information available to the public regarding COVID-19 makes it difficult to delineate accurate information from inaccurate information [44]. Practitioners have a preexisting rapport with their patients that they should use to shift the balance toward accurate information. This patient-provider relationship may engender trust that does not exist with larger health or government entities. Practitioners should capitalize on this rapport to convey accurate, timely information regarding risk factors, protective measures, coping techniques, financial relief, and food banks.

Policy makers should encourage growth in areas of mental health support that are most feasible during this time. Telemental health, for example, has been shown to be highly effective, cost-efficient, and accessible, especially in isolated communities [45]. Online mental health assessments and self-directed mental health interventions have also been widely introduced in China, with their effectiveness remaining to be seen [46].

Future research should continue to track the mental health effects of the pandemic as it progresses. There may be future waves of illness that impact social distancing recommendations and requirements. These, in turn, may impact mental health. Longitudinal investigation of these effects is recommended. Future studies should make concerted efforts to obtain a representative sample. Representative state-specific samples are available through various entities for a fee. In addition, specific outreach to underrepresented populations is recommended. Knowledge of these fluctuations in population mental health can be used by public health practitioners, mental health practitioners, and policy makers in their decision making and in their framing of recommendations.

Acknowledgments

This study was supported by NIH/NCRR Colorado CTSI Grant Number UL1 RR025780. Its contents are the authors' sole responsibility and do not necessarily represent official National Institutes of Health (NIH) views.

Conflicts of Interest

None declared.

References

1. Johns Hopkins. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE). Coronavirus Resource Center. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-10-15]
2. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
3. Lima CKT, Carvalho PMDM, Lima IDAAS, Nunes JVADO, Saraiva JS, de Souza RI, et al. The emotional impact of Coronavirus 2019-nCoV (new Coronavirus disease). Psychiatry Res 2020 May;287:112915 [FREE Full text] [doi: 10.1016/j.psychres.2020.112915] [Medline: 32199182]
4. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/nejmp2008017]
5. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. Int J Soc Psychiatry 2020 Jun 31;66(4):317-320. [doi: 10.1177/0020764020915212] [Medline: 32233719]
6. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during covid-19 pandemic. BMJ 2020 Mar 26;368:m1211. [doi: 10.1136/bmj.m1211] [Medline: 32217624]
7. Cénat JM, Mukunzi JN, Noorishad P, Rousseau C, Derivois D, Bukaka J. A systematic review of mental health programs among populations affected by the Ebola virus disease. J Psychosom Res 2020 Feb 13;131:109966. [doi: 10.1016/j.jpsychores.2020.109966] [Medline: 32087433]
8. Paladino L, Sharpe R, Galwankar S, Sholevar F, Marchionni C, Papadimos T, American College of Academic International Medicine (ACAIM). Reflections on the Ebola Public Health Emergency of International Concern, Part 2: The Unseen

Epidemic of Posttraumatic Stress among Health-care Personnel and Survivors of the 2014-2016 Ebola Outbreak. J Glob Infect Dis 2017;9(2):45-50 [FREE Full text] [doi: 10.4103/jgid.jgid_24_17] [Medline: 28584454]
9. Bo H, Li W, Yang Y, Wang Y, Zhang Q, Cheung T, et al. Posttraumatic stress symptoms and attitude toward crisis mental health services among clinically stable patients with COVID-19 in China. Psychol Med 2020 Mar 27:1-2. [doi: 10.1017/s0033291720000999]
10. Pacella ML, Armelie A, Boarts J, Wagner G, Jones T, Feeny N, et al. The impact of prolonged exposure on PTSD symptoms and associated psychopathology in people living with HIV: a randomized test of concept. AIDS Behav 2012 Jul 20;16(5):1327-1340 [FREE Full text] [doi: 10.1007/s10461-011-0076-y] [Medline: 22012149]
11. Penzak S, Reddy YS, Grimsley SR. Depression in patients with HIV infection. Am J Health Syst Pharm 2000 Feb 15;57(4):376-386. [doi: 10.1093/ajhp/57.4.376] [Medline: 10714976]
12. Mak IWC, Chu CM, Pan PC, Yiu MGC, Chan VL. Long-term psychiatric morbidities among SARS survivors. Gen Hosp Psychiatry 2009 Jul;31(4):318-326 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.03.001] [Medline: 19555791]
13. Maunder RG. Was SARS a mental health catastrophe? Gen Hosp Psychiatry 2009 Jul;31(4):316-317 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.04.004] [Medline: 19555790]
14. Wu KK, Chan SK, Ma TM. Posttraumatic stress after SARS. Emerg Infect Dis 2005 Aug;11(8):1297-1300 [FREE Full text] [doi: 10.3201/eid1108.041083] [Medline: 16102324]
15. US Bureau of Labor Statistics. URL: https://www.bls.gov [accessed 2020-10-15]
16. Avčin BA, Kučina AU, Sarotar BN, Radovanović M, Plesničar BK. The present global financial and economic crisis poses an additional risk factor for mental health problems on the employees. Psychiatr Danub 2011 Sep;23 Suppl 1:S142-S148. [Medline: 21894123]
17. Mucci N, Giorgi G, Roncaioli M, Fiz Perez J, Arcangeli G. The correlation between stress and economic crisis: a systematic review. NDT 2016 Apr:983. [doi: 10.2147/ndt.s98525]
18. Cacioppo JT, Cacioppo S. Social Relationships and Health: The Toxic Effects of Perceived Social Isolation. Soc Personal Psychol Compass 2014 Feb 01;8(2):58-72 [FREE Full text] [doi: 10.1111/spc3.12087] [Medline: 24839458]
19. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar 11;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
20. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. Heart 2016 Jul 01;102(13):1009-1016. [doi: 10.1136/heartjnl-2015-308790] [Medline: 27091846]
21. Wang J, Lloyd-Evans B, Giacco D, Forsyth R, Nebo C, Mann F, et al. Social isolation in mental health: a conceptual and methodological review. Soc Psychiatry Psychiatr Epidemiol 2017 Dec 28;52(12):1451-1461 [FREE Full text] [doi: 10.1007/s00127-017-1446-1] [Medline: 29080941]
22. Leigh-Hunt N, Bagguley D, Bash K, Turner V, Turnbull S, Valtorta N, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health 2017 Nov;152:157-171. [doi: 10.1016/j.puhe.2017.07.035] [Medline: 28915435]
23. Hamel L, Lopes L, Muñana C, Kates J, Michaud J, Brodie M. KFF Coronavirus Poll: March 2020. Kaiser Family Foundation. 2020 Mar 17. URL: https://www.kff.org/coronavirus-covid-19/poll-finding/kff-coronavirus-poll-march-2020/ [accessed 2020-10-15]
24. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
25. Twenge J, Joiner E. Mental stress among US Adults During the COVID-19 pandemic. PsyArXiv 2020:1. [doi: 10.31234/osf.io/wc8ud]
26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
27. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5(1):63. [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
28. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010 Jul 19;8(4):348-353. [doi: 10.1370/afm.1139] [Medline: 20644190]
29. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008 Sep;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
30. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
31. Elo A, Leppänen A, Jahkola A. Validity of a single-item measure of stress symptoms. Scand J Work Environ Health 2003 Dec;29(6):444-451 [FREE Full text] [doi: 10.5271/sjweh.752] [Medline: 14712852]

32. National Institute of Mental Health. Prevalence of Major Depressive Episode Among Adults. 2019 Feb 20. URL: https://www.nimh.nih.gov/health/statistics/major-depression.shtml [accessed 2020-06-15]
33. Prevalence of Any Anxiety Disorder Among Adults. National Institute of Mental Health. 2017. URL: https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder.shtml [accessed 2020-06-15]
34. Lloyd K, Devine P. Psychometric properties of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in Northern Ireland. J Ment Health 2012 Jun 10;21(3):257-263. [doi: 10.3109/09638237.2012.670883] [Medline: 22574955]
35. Yang Y, Li W, Zhang Q, Zhang L, Cheung T, Xiang Y. Mental health services for older adults in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e19. [doi: 10.1016/s2215-0366(20)30079-1]
36. Modini M, Joyce S, Mykletun A, Christensen H, Bryant RA, Mitchell PB, et al. The mental health benefits of employment: Results of a systematic meta-review. Australas Psychiatry 2016 Aug 10;24(4):331-336. [doi: 10.1177/1039856215618523] [Medline: 26773063]
37. Moscone F, Tosetti E, Vittadini G. The impact of precarious employment on mental health: The case of Italy. Soc Sci Med 2016 Jun;158:86-95. [doi: 10.1016/j.socscimed.2016.03.008] [Medline: 27115334]
38. Baicker K, Allen HL, Wright BJ, Taubman SL, Finkelstein AN. The Effect of Medicaid on Management of Depression: Evidence From the Oregon Health Insurance Experiment. Milbank Q 2018 Mar 05;96(1):29-56 [FREE Full text] [doi: 10.1111/1468-0009.12311] [Medline: 29504203]
39. Gender and Stress. American Psychological Association. 2012. URL: https://www.apa.org/news/press/releases/stress/2010/gender-stress#:~:text=Women%20are%20more%20likely%20than%20men%20(28%20percent%20vs.,10%20(39%20percent)%20men [accessed 2020-06-17]
40. Handa RJ, Chung WCJ. Chapter 14 - Gender and Stress. In: Fink G, editor. Stress: Physiology, Biochemistry, and Pathology. Amsterdam, Netherlands: Elsevier; 2019:165-176.
41. Jessop DC, Reid M, Solomon L. Financial concern predicts deteriorations in mental and physical health among university students. Psychol Health 2020 Feb 10;35(2):196-209. [doi: 10.1080/08870446.2019.1626393] [Medline: 31181966]
42. Jones AD. Food Insecurity and Mental Health Status: A Global Analysis of 149 Countries. Am J Prev Med 2017 Aug;53(2):264-273 [FREE Full text] [doi: 10.1016/j.amepre.2017.04.008] [Medline: 28457747]
43. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/nejmp2008017]
44. Mitchell A, Gottfried J, Barthel M, Sumida N. Distinguishing Between Factual and Opinion Statements in the News. Pew Research Center. 2018 Jun 18. URL: https://www.journalism.org/2018/06/18/distinguishing-between-factual-and-opinion-statements-in-the-news/ [accessed 2020-10-15]
45. Langarizadeh M, Tabatabaei M, Tavakol K, Naghipour M, Rostami A, Moghbeli F. Telemental Health Care, an Effective Alternative to Conventional Mental Care: a Systematic Review. Acta Inform Med 2017 Dec;25(4):240-246 [FREE Full text] [doi: 10.5455/aim.2017.25.240-246] [Medline: 29284913]
46. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e17-e18. [doi: 10.1016/s2215-0366(20)30077-8]

Abbreviations

GAD-7: Generalized Anxiety Disorder 7-item
PHQ-2: Patient Health Questionnaire-2
REDCap: Research Electronic Data Capture
SWEMWBS: Short Warwick-Edinburgh Mental Well-being Scale

Edited by A Tal; submitted 02.07.20; peer-reviewed by D Govindasamy, R Nogueira-Arjona, JM Chen; comments to author 24.07.20; revised version received 11.08.20; accepted 28.08.20; published 23.10.20.

Please cite as:
Jewell JS, Farewell CV, Welton-Mitchell C, Lee-Winn A, Walls J, Leiferman JA
Mental Health During the COVID-19 Pandemic in the United States: Online Survey
JMIR Form Res 2020;4(10):e22043
URL: http://formative.jmir.org/2020/10/e22043/
doi:10.2196/22043
PMID:33006939

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org