Original Papers

Deep Learning With Anaphora Resolution for the Detection of Tweeters With Depression: Algorithm Development and Validation Study (e19824)
Akkapon Wongkoblap, Miguel Vadillo, Vasa Curcin. ............................................................. 2

A Compassion-Focused Ecological Momentary Intervention for Enhancing Resilience in Help-Seeking Youth: Uncontrolled Pilot Study (e25650)
Christian Rauschenberg, Benjamin Boecking, Isabell Paetzold, Koen Schruers, Anita Schick, Thérèse van Amelsvoort, Ulrich Reininghaus. . . 1

Mental Health and the Perceived Usability of Digital Mental Health Tools Among Essential Workers and People Unemployed Due to COVID-19: Cross-sectional Survey Study (e28360)
Felicia Mata-Greve, Morgan Johnson, Michael Pullmann, Emily Friedman, Isabell Griffith Fillipo, Katherine Comtois, Patricia Arean. .................. 36
Deep Learning With Anaphora Resolution for the Detection of Tweeters With Depression: Algorithm Development and Validation Study

Akkapon Wongkoblap¹,²,³*, MSc; Miguel A Vadillo⁴,⁵*, PhD; Vasa Curcin¹,⁴*, PhD

¹Department of Informatics, King's College London, London, United Kingdom
²DIGITECH, Suranaree University of Technology, Nakhon Ratchasima, Thailand
³School of Information Technology, Suranaree University of Technology, Nakhon Ratchasima, Thailand
⁴School of Population Health and Environmental Sciences, King's College London, London, United Kingdom
⁵Departamento de Psicología Básica, Universidad Autónoma de Madrid, Madrid, Spain

*all authors contributed equally

Corresponding Author:
Akkapon Wongkoblap, MSc
DIGITECH
Suranaree University of Technology
111 University Avenue, Muang
Nakhon Ratchasima, 30000
Thailand
Phone: 66 44224336
Email: wongkoblap@sut.ac.th

Abstract

Background: Mental health problems are widely recognized as a major public health challenge worldwide. This concern highlights the need to develop effective tools for detecting mental health disorders in the population. Social networks are a promising source of data wherein patients publish rich personal information that can be mined to extract valuable psychological cues; however, these data come with their own set of challenges, such as the need to disambiguate between statements about oneself and third parties. Traditionally, natural language processing techniques for social media have looked at text classifiers and user classification models separately, hence presenting a challenge for researchers who want to combine text sentiment and user sentiment analysis.

Objective: The objective of this study is to develop a predictive model that can detect users with depression from Twitter posts and instantly identify textual content associated with mental health topics. The model can also address the problem of anaphoric resolution and highlight anaphoric interpretations.

Methods: We retrieved the data set from Twitter by using a regular expression or stream of real-time tweets comprising 3682 users, of which 1983 self-declared their depression and 1699 declared no depression. Two multiple instance learning models were developed—one with and one without an anaphoric resolution encoder—to identify users with depression and highlight posts related to the mental health of the author. Several previously published models were applied to our data set, and their performance was compared with that of our models.

Results: The maximum accuracy, F1 score, and area under the curve of our anaphoric resolution model were 92%, 92%, and 90%, respectively. The model outperformed alternative predictive models, which ranged from classical machine learning models to deep learning models.

Conclusions: Our model with anaphoric resolution shows promising results when compared with other predictive models and provides valuable insights into textual content that is relevant to the mental health of the tweeter.

(JMIR Ment Health 2021;8(8):e19824) doi:10.2196/19824

KEYWORDS

depression; mental health; Twitter; social media; deep learning; anaphora resolution; multiple-instance learning; depression markers
Introduction

Background
Mental health problems are widely recognized as major public health challenges worldwide. According to the World Health Organization, 264 million people were affected by depression globally in 2020 [1]. Mental illness, in general, is one of the leading causes of the global burden of this disease. It was estimated that in England, 105 billion British pounds (US $145 billion) were spent on mental health services and treatments or lost in productivity at work in 2018 [2], with the global costs expected to rise to US $6 trillion by 2030 [3]. A significant contributor to this cost is that people living with mental health problems sometimes receive inaccurate assessments [1]. This highlights the need for effective mental health services and a novel approach for diagnosing mental health disorders.

User-generated content on social media, reviews, blogs, and message board platforms offers an opportunity for researchers to explore and classify the huge amount of content in different domains, such as marketing [4], politics [5], and health [6-8], thereby providing a rapid method to understand user-created text and expressed emotion using text classification algorithms. Social networking (eg, Facebook and LinkedIn) and microblogging platforms (eg, Twitter and Tumblr) provide internet users with a safe space to post their feelings, thoughts, and activities. With some users publicly expressing their mental health statuses on their profiles, it becomes possible to train classification engines to detect internet users with mental health problems [9,10]. Using Twitter data, in particular, studies have examined users with depression [11-14], postpartum depression [15], anxiety, obsessive compulsive disorder, and posttraumatic stress disorder [11,16]. In addition, Facebook data were also used to detect users with depression [17,18] and postpartum depression [19].

Generally, text classifiers and user classification models tend to be developed separately. This presents a challenge for researchers who want to simultaneously understand both text sentiment analysis and user sentiment analysis. In this paper, we present a predictive model that can detect users with depression and identify their tweets as those related to health. An ideal technique for developing this type of model is multiple instance learning (MIL) [20], where the model can learn from a set of labeled bags or users instead of a set of individual instances or user-generated messages.

Anaphora resolution is an established natural language processing (NLP) problem and an emerging field in the analysis of social media content that helps with determining which previously mentioned person is the subject of a subsequent statement and understanding references to someone in the content on social media. This is particularly relevant to social media, as posts may frequently refer to individuals other than a third party.

The objective of this study is to investigate whether user-generated content from Twitter can be used to detect users with depression. This raises three research questions:

1. Can MIL be used to develop a predictive model for detecting users with depression from their tweets?
2. Can sentiments of unlabeled tweets be predicted from the labels of users with depression?
3. Can anaphora resolution be combined with MIL to eliminate false positives?

This study focuses on text analysis, predictive models for detecting social network users with mental disorders, and MIL. The most relevant studies published to date are reviewed below.

Text analysis is an NLP approach for identifying information within text. This technique has been developed to understand the textual content automatically and computationally. During the early stages of sentiment and emotion analysis, researchers manually annotated the text [22]. With the possibility of identifying emotions in text, the content has been computationally analyzed using a keyword or corpus-based approach and a learning-based approach [23,24].

The learning-based approach uses a predictive model to determine the relationship between an input and output word. Word embedding is a common learning-based technique that transforms the words of a document into dimensional vectors for word representation and determines word similarity. Global Vectors for Word Representation (GloVe) is a word-embedding approach that computes and aggregates word co-occurrence for representing the closest linguistic or semantic similarity between co-occurrent words as vectors [25]. GloVe was trained on several textual data sets, such as Wikipedia and common crawl (a copy of web content), and supported 50D, 100D, 200D, and 300D vectors.

Anaphora resolution is another text analysis problem related to determining which person is mentioned within textual content. There are three reference resolution algorithms [26]. The rule-based entity resolution extracts syntactic rules and semantic knowledge from the text. The statistical and machine learning–based entity resolution is a method to understand the coreference of a reference to an early entity. Deep learning for entity resolution reduces handcrafted feature requirements and represents words as vectors conveying semantic units. Aktaş et al [21] investigated anaphora resolution for conversations on
De Choudhury and Gamon [13] pioneered NLP and machine learning approaches for developing predictive models to detect users with mental disorders from social network data using a mental health screening questionnaire and linguistic analysis tools to extract emotional words and web-based behaviors from users’ posts. However, the screening and data collection process was time consuming, and Coppersmith et al [11] introduced an automatic data gathering method using keywords to find the target users and programmatically retrieve the posts.

Following these initial studies, a number of novel methods have emerged for predicting mental disorders in social network users. The early work focused on classical supervised machine learning techniques and traditional text analysis approaches.

The psychometric analysis of textual content was used to compute the percentage of emotional, functional, and social concern words [13,15]. Linguistic inquiry and word count (LIWC) was used to compute the percentage of words relevant to categories from each tweet. The extracted percentages were then used to train a predictive model based on a support vector machine with a radial basis function [13].

Language models have been applied to analyze social media texts to address spelling errors, shortenings, and emoticons [11]. The language model was developed from an n-gram, which learns from the sequences of text and computes the probability of unseen text relevant to a category of the trained model. This model scored the probabilities of users with depression based on a higher probability of the positive class language model trained from the tweets of users with depression or the negative class language model developed from the tweets of control users [11].

A predictive model based on topic models was developed from the social network profiles of clinically diagnosed patients [17]. The topic model used latent Dirichlet allocation to extract topics from the text. All tweets from each user were used to compute 200 topics, which were then used to develop a logistic regression model for classifying the users with depression [17].

Building on the popularity of neural networks, novel models have been developed using word embedding [27,28] and deep neural network models [28]. The Usr2Vec model transformed text into an embedding matrix, where words commonly used together were represented in closely dimensional spaces for classifying users. The embeddings were learned from users’ tweets and then summarized as user representations. The embedding matrices were used to train a predictive model using a multinomial logistic regression technique [27].

The deep learning model uses word embeddings to represent the sequential words of users’ tweets. A predictive model was trained using a 1D convolutional neural network (CNN) and a global max pooling layer [28].

In addition to the textual content of the posts, a number of writing features can be analyzed: post or blog lengths, time gap between consecutive posts, and day of the week and time of the day of postings. Further network features of interest include likes, numbers of followers or following, characteristics of comments on other users’ posts compared with original posts, and numbers of shares or retweets. Image analysis was used to characterize user posts [29,30].

To develop a predictive model, this study focused on MIL. It is a supervised learning technique first proposed by Keeler et al [31,32]. Although classical supervised learning requires an instance and a single label to learn during the training process, MIL can learn from a bag of instances \(X=x_1, x_2, \ldots, x_N\). Each instance \(x_n\) can be independent and has its own individual label, \(y_n\), where \(y_n \in \{0, 1\}\) for \(n=1, \ldots, N\), and it is assumed that each \(y_n\) is unknown during the training process. On the basis of these assumptions, an MIL classifier can predict a label \(Y\) for a given bag \(X\) as follows:

\[
Y = \begin{cases} 
1 & \text{if} \ 	ext{there is at least one positive instance} \\
0 & \text{otherwise}
\end{cases}
\]

On the basis of these assumptions, MIL can provide an extreme result \(Y=1\) in the case of having a predicted positive-instance label \(y_n=1\) in a given input \(X\). The relaxation of the MIL assumption can be computed using aggregated probabilistic distributions of instances, where \(Y=P(x_n)\) for \(n=1, \ldots, N\).

The purpose of MIL is to facilitate the development of a predictive model for detecting social media users with depression and instantly label each of the posts associated with either mental health or other topics. Normally, data sets from social networking are labeled at the user level but not at the post level. This makes it difficult to find a change in patterns in the message topics posted on social networks.

MIL models have been widely applied to image classification [32], object detection [33], image annotation [34], medical image and video analysis [35,36], sentence selection [37], and document classification [38]. In document sentiment analysis, Angelidis and Lapata [20] proposed the MIL network (MILNET) to classify web-based review documents and instantly identify the sentiment polarity of each segment of given documents. MILNET comprises segment encoding, segment classification, and document classification via an attention mechanism. Segmented encoding transformed sentences in a document into segments via word-embedding matrices and a CNN. Each segment representation was classified using a softmax classifier. An attention mechanism based on a bidirectional gated recurrent unit (GRU) was used to weight the important segments to make a final document prediction as the weighted sum of the segment distributions. MILNET performed well in predicting the sentiment of a document and identifying the sentiment of the text segments but was not as successful in identifying a person mentioned in the document.

In this study, we adopt the MIL approach to develop two models, namely multiple instance learning for social network (MIL-SocNet) and multiple instance learning with an anaphoric resolution for social network (MILA-SocNet), to classify users with depression and highlight published posts associated with the mental health topic of a tweeter. Both models use novel document segment encoding, a tweet encoder, and user
representation rather than a document vector. The latter model also includes the anaphora resolution, which further improves the performance.

**Methods**

**Data Set**

The data set was retrieved from Twitter, which provides an application programming interface (API) to search public tweets using regular expressions or stream real-time tweets. This study collected only tweets and users set as public. All collected tweets and users were anonymized. This study was approved by the King’s College Research Ethics Committee (reference number LRS-16/17-4705).

We selected a group of users with depression using the method proposed by Coppersmith et al [11]. Specifically, a regular expression was used to search tweets that contained the statement “I was diagnosed with depression” between January and May 2019. This resulted in 4892 tweets from 4545 unique users, who were then manually screened to ensure that the tweets did not refer to jokes, quotes, or someone else’s depression symptoms. After removing these messages, all tweets in the profiles of the users who posted the tweets were downloaded. After verification, 2132 unique users were included in this data set.

A control group was randomly selected from a list of 2036 users who posted tweets between June 1 and June 7, 2019. Users from the group with depression were removed from the list of the control group.

The limits imposed by the Twitter API allowed us to only download the 3200 most recent tweets of all verified users from the depressed and control groups. In total, 5 million tweets were collected from the 2132 users with depression and 4.2 million tweets from the 2036 users with no declared depression.

**Preprocessing**

Before developing our MIL model, several transformations were performed on the data set. First, the user ID in each tweet was replaced by a generic user. Similarly, any numbers mentioned in tweets were replaced by the number and any specific URLs by url. The # character in each hashtag was replaced by the string hashtag (eg, #depression became hashtag depression). Finally, users with fewer than 100 tweets or less than 80% of tweets in English were removed from the data set, resulting in 3682 users, 1983 with declared depression and 1699 with no declared depression, as depicted in on the left-hand side of Figure 1. In addition, other dimensions of the data set were explored, as shown in Figure 1. Figure 2 illustrates the distribution of the number of tweets between the depressed and control groups. Slight differences were present between the control and depressed groups.

All tweets in our final data set were embedded from pretrained GloVe word vectors. GloVe is an unsupervised machine learning approach and an NLP technique that represents a word as a set of word vectors. GloVe computes and aggregates word co-occurrences to create a vector representation of the closest linguistic or semantic similarity between co-occurrence words [22]. As explained earlier, GloVe was trained on several textual data sets, for example, Wikipedia and common crawl (a copy of web content), and supported 50D, 100D, 200D, and 300D vectors. However, our study used pretrained word vectors trained on 2 billion tweets and 100D vectors to transform our tweets into word embedding.

**Predictive Model**

**Overview**

This section describes the structure of our predictive model to classify a Twitter user with depression. This section will explain how an MIL model with supervised neural networks classifies users and provides a changing pattern of generated text associated with mental health or other topics.

Our proposed MIL-SocNet architecture comprises a tweet encoder, word attention on a tweet, tweet classification, a user encoder, tweet attention, and user classification (Figure 3). The differences between MIL-SocNet and the basic MILNET architecture are the tweet encoder and word attention, respectively. Our model uses a GRU, whereas MILNET uses a CNN and does not have an attentional mechanism.

Furthermore, the MIL-SocNet model was extended with an anaphoric resolution to create the MILA-SocNet model. We present this model to improve performance by adding an anaphora resolution encoder to ensure that the algorithm focuses on posts related to the author (Figure 4).

**Tweet Encoder**

The first layer of our proposed model transforms each tweet into a machine-readable form. First, tweets were transformed into word-embedding matrices. Each user publishes \( j = 1, 2, \ldots, n \) tweets, where \( n \) is the number of tweets used to train the model. Each tweet contains \( k = 1, 2, \ldots, i \) words, where \( i \) is the number of words in each tweet and varies from post to post. \( W_jk \) represents the \( k \)th word in the \( j \)th tweet. Every \( w_{jk} \) is then embedded through an embedding matrix \( W \) to be received a word vector \( x_{jk} \). This layer embeds all words \( w_{jk} \) of \( j \)th post to the word vector:

\[
x_{jk} = w_{jk}W_{ewj} \quad j \in [1, n] \text{ and } k \in [1, i]
\]

The abovementioned equation operates \( i \) times. After embedding all words, a bidirectional GRU is used to encode the vector:

\[
\begin{align*}
x_{jk-1} & = x_{jk-1} + x_{jk} \\
x_{jk+1} & = x_{jk+1} + x_{jk}
\end{align*}
\]

The bidirectional GRU presents a hidden representation of \( h_{jk} \), which is concatenated from \( x_{jk-1} \) and \( x_{jk+1} \). The word hidden vector \( h_{jk} \) is then sent to an attention mechanism to select the important words.
Word Attention on a Tweet

Not every word equally represents tweet meanings. An attention mechanism is used to select words that best capture the relevant meaning of a tweet. The attention layer comprises a tanh function to produce an attention vector $a_{jk}$ of the $k$th word in the $j$th tweet, where $a_w$ and $b_w$ are weights and bias, respectively.

$$u_{jk} = \tanh (W_u h_{jk} + b_u)$$

The importance of words or attention weights $a_{jk}$ is calculated via the normalized similarity of $u_{jk}$ with the context vector of the word level $u_w$, which is learned and updated during the training step.

Finally, the tweet vector $t_j$ is computed using the weighted sum of word importance with the hidden representation of $h_{jk}$ generated from the bidirectional GRU.

Tweet Classification

To make a prediction about a tweet related to either a mental health or another topic, each tweet vector $t_1, t_2, ..., t_n$ from the attention layer is classified through a softmax function [39].

The function generates the probabilities of tweet labels $p$, where $p$ with 1 denoting a mental health–related post and 0 denoting a non–mental health–related post. The labels used to train this layer are derived and computed from the labels of the user level only. The parameters $W_c$ and $b_c$ are learned and updated during the training step. Every predicted tweet label is used to teach a predictive model and detect a user with depression.

User Encoder

Detecting users with depression requires a pattern to differentiate between user groups. To predict these users, this study used a temporal pattern of posting generated from the tweet classification layer. This layer concatenates the probabilities of every classified tweet label into a single list of label probabilities called user representation. The user representations between the 2 groups are expected to differ, which will be explored and illustrated in the Discussion section. Then, user representation is passed through a bidirectional GRU to learn the changing patterns of text categories over the observation time. This generates the forward hidden state $h_f$ and the backward hidden state $h_b$ of the user representation. Finally, they were concatenated to $h_j$.

Anaphora Resolution Encoder

For the MILA-SocNet model with anaphora resolution, pronoun features from LIWC [40] are used to add informative interpretations to each tweet. Every tweet was analyzed for emotions, thinking styles, social states, parts of speech, and psychological dimensions.

Each tweet is combined between the extracted pronoun features $s_j$ from the LIWC and a tweet classified label $p_j$ from the tweet classification layer, where $p$ with $p$ represents the extracted features in the $j$th tweet. This yields the following anaphora resolution vector:

The vector is then passed through a bidirectional GRU to learn the text category and anaphoric features. This generates $h_j$ combined from the forward hidden state $h_f$ and backward hidden state $h_b$.

Tweet Attention

Not all user tweets were equally associated with depression. Some tweets may contain cues relevant to depression, whereas others may not. For this purpose, an attention mechanism is applied to reward tweets that correctly represent the characteristics and are important for correctly detecting a user with depression. This layer performs similarly in both MIL–SocNet and MILA–SocNet. A multilayer perceptron (MLP) produces the attention vector $u_{j}$ of the $j$th tweet. The parameter $W_t$ denotes the weights of the tweet and parameter $b_t$ represents the bias of the tweet.

The attention weights of tweets or important tweets $\alpha_j$ are computed through the similarity of $u_j$ with the context vector of tweet level $u_c$, which is learned and updated during the training step.

The user vector $v$ is achieved by summarizing all the information of the tweet label possibilities of a user.

User Classification

Finally, a predictive model for detecting a user with depression can be achieved through the user vector $v$ derived from encoding...
the concatenation of the probabilities and the attention weights of the classified tweet labels from the user. A softmax function was again used to perform the classification.

Training the MIL Model

To train MILA-SocNet and MIL-SocNet, we used the Keras library with TensorFlow backend, a Python library for neural network APIs. We used an adaptive and momental bound method (AdaMod) [41], and the binary cross-entropy loss function to minimize loss. Every tweet from each user was tokenized and limited to 55 tokens or words. The model was trained using 2000 recent tweets from each user, with users with fewer than 2000 tweets having empty tweets padded with 0 values to achieve the matching length. To eliminate overfitting, dropout and early stopping were applied to the model during the training step.

Both our models and replicated models were trained and tested with holdout cross-validation. We split the users experiencing depression into four equal chunks and trained the models against all control users. Thus, each round used 496 users experiencing depression (22.60%) and 1699 control users (77.40%), mirroring the real-world incidence of depression. From the total users included in each round, 20% were used as test sets to evaluate the performance of the models. To reserve the same proportions of classes between the training and test tests, stratified cross-validation was used. Figure 5 shows the cross-validation process.

Model Evaluation

To predict whether each Twitter user was likely to be depressed, we also trained a set of published predictive models ranging from classical machine learning to deep learning techniques by using user-generated textual content. Accuracy, precision, recall, and F1 score were averaged across the test sets. Each model was trained and tested with the same samples in each round; however, data transformations differed in some cases, as explained in the Background section.

To compute the predictive performance of models for detecting social network users with depression, we used the following metrics:

To further compare the performance of MILA-SocNet with the other published models, Akaike information criterion (AIC) was applied across all the models. AIC is a commonly used tool for model comparison and selection [42,43] that measures the information loss in each model, considering the model’s complexity as well. AIC is defined as follows:

where $n$ is the number of samples and $K$ is the number of parameters or features of a model. $\ln$ denotes the natural logarithm of likelihood [44]. The equation also uses bias adjustment because of the small sample size [45,46]. A lower AIC value indicates better performance.

Results

This section shows the performance of MILA-SocNet and MIL-SocNet and compares their results in terms of accuracy, precision, recall, and F1 score against several published models including LIWC [13], language [11], topic [17], Usr2Vec [27], and deep learning [28] models, as explained in the Background section.

Table 1 shows the performance of our proposed MILA-SocNet and MIL-SocNet models against the alternative models. As observed, the MILA-SocNet achieves a maximum accuracy (92%), precision (92%), recall (92%), and F1 score (92%), immediately followed by the MIL-SocNet. The MIL-SocNet yielded an accuracy, precision, recall, and F1 score of 90%, 91%, 90%, and 90%, respectively. Each model was evaluated using the area under the curve of the receiver operating characteristic curve. As can be seen in Figure 6, the MILA-SocNet and MIL-SocNet models achieved the highest areas under the curve—93% in both cases. It should be noted that in those studies, the replicated models were reported with different proportions of classes. These results might be higher or lower than our reported results. In our study, the baseline result was 77% in the case of predicting the majority class in all cases. As can be observed, all the models achieved results that were above this baseline.

Table 2 lists the AIC values for each model. The likelihood was computed from the model-based probabilities of the observed labels. The number of parameters of the MILA-SocNet, MIL-SocNet, and deep learning models were recovered from the number of trainable parameters reported by the Keras library. The number of parameters of the language model was taken from the number of vocabularies in the positive and negative language models. The number of parameters of LIWC, Usr2Vec, and topic models were features in the models. The likelihoods and AICs were averaged from cross-validation, as explained earlier. As can be observed, MILA-SocNet achieves the lowest AIC, reflecting the best performance.
Table 1. Performance of our proposed MILA-SocNet (multiple instance learning with an anaphoric resolution for social network) and MIL-SocNet (multiple instance learning for social network) models and all replicated models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy, %</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MILA-SocNet</td>
<td>92.14</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>MIL-SocNet</td>
<td>90.49</td>
<td>0.91</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Deep learning</td>
<td>89.07</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>Usr2Vec</td>
<td>84.38</td>
<td>0.84</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>LIWC</td>
<td>83.31</td>
<td>0.83</td>
<td>0.83</td>
<td>0.81</td>
</tr>
<tr>
<td>Language</td>
<td>81.61</td>
<td>0.80</td>
<td>0.82</td>
<td>0.79</td>
</tr>
<tr>
<td>Topic</td>
<td>80.13</td>
<td>0.78</td>
<td>0.80</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 2. The Akaike information criterion (AIC) results against all models. Each row is reported with the number of parameters (K), the residual sum of squares, and the AIC. A lower AIC is better.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of parameters, K</th>
<th>Likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MILA-SocNet</td>
<td>59,668</td>
<td>−143.72</td>
<td>−597.05</td>
</tr>
<tr>
<td>MIL-SocNet</td>
<td>56,296</td>
<td>−210.22</td>
<td>−464.45</td>
</tr>
<tr>
<td>Deep learning</td>
<td>138,502</td>
<td>−309.97</td>
<td>−260.84</td>
</tr>
<tr>
<td>Language</td>
<td>16695.5</td>
<td>−420.31</td>
<td>−61.03</td>
</tr>
<tr>
<td>LIWC</td>
<td>93</td>
<td>−169.62</td>
<td>575.92</td>
</tr>
<tr>
<td>Usr2Vec</td>
<td>100</td>
<td>−190.28</td>
<td>640.32</td>
</tr>
<tr>
<td>Topic</td>
<td>200</td>
<td>−276.42</td>
<td>1290.66</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

In this study, we presented two novel MIL models for detecting social network users with depression based on their self-identifying tweets. The original MIL-SocNet model was extended with anaphoric resolution to produce the second MILA-SocNet model. We also compared the performance of both models with that of several previously published models. As can be seen from Tables 1 and 2, MILA-SocNet and MIL-SocNet outperformed all other models in all metrics. We now look at several potential reasons for this result.

Although deep learning models can be trained on raw textual data, traditional machine learning models (eg, the LIWC, language, topic, and Usr2Vec models) require feature extraction to be performed using external tools, which may introduce the additional risk of losing useful information from short textual data [47,48]. For instance, misspelled and abbreviated words in tweets may not be present in the dictionary of an extraction tool, resulting in the mislabeling of words. This may be one of the reasons why traditional machine learning techniques performed worse than our proposed models.

Another reason for the performance gap may be that the sequential ordering of words in a tweet and tweets posted on a timeline may influence model performance. Training a predictive model with traditional machine learning methods requires aggregated data, which may cause the loss of contextual information compared with deep neural networks that can learn from the sequential information in the data [49-52].

Unlike the deep learning model that we have compared against [28], MILA-SocNet and MIL-SocNet used an attention mechanism that highlights words and tweets relevant to mental health. This attention mechanism may have contributed to our proposed models outperforming the deep learning model, even though our approach is also based on deep learning techniques.

Another important point to consider is that the addition of anaphoric resolution improves the performance of the base MIL model. The difference between MILA-SocNet and MIL-SocNet is only in anaphora resolution encoding, which highlights posts related to the tweeters rather than someone else. This is an important feature that has not been widely investigated in the field and should be considered while designing future studies.

We further explored our proposed models by comparing the model performance under different conditions. A set of different parameters was used to train the models. The number of each user’s posts used to train a model ranged from 500 to 3200 posts. The numbers of embedded dimensions were 50 and 100. The lengths of word tokens used to train the models were 18...
and 55 tokens, respectively. Table 3 and Figure 7 show the predictive results of MILA-SocNet and MIL-SocNet with different parameters. Longer post length and longer word token provide better results, which is expected as these provide more textual content. Furthermore, models with fewer embedded dimensions perform worse than models with more dimensions. After training the models, we investigated their interpretability by observing the attention weights to find out which tweets the model paid most attention to. Two users from each group were randomly selected from those correctly labeled by our model, and attention weights were extracted from the tweet attention layer. Textbox 1 highlights the tweets that achieved the highest and lowest weights for these 4 users, offering some insight into model's decision-making. Our predictive model with anaphoric resolution can identify tweets related to the tweeters’ own experiences.

Table 3. Performance of MILA-SocNet (multiple instance learning with an anaphoric resolution for social network) and MIL-SocNet (multiple instance learning for social network) with different parameters. The first number in the model name (first column) represents the number of posts, the second is the number of embedded dimensions, and the last is the number of word tokens.

<table>
<thead>
<tr>
<th>Model name</th>
<th>MILA-SocNet models</th>
<th>MIL-SocNet models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy, %</td>
<td>Precision</td>
</tr>
<tr>
<td>2000-100-55</td>
<td>92.14</td>
<td>0.92</td>
</tr>
<tr>
<td>500-100-55</td>
<td>85.88</td>
<td>0.86</td>
</tr>
<tr>
<td>3200-100-18</td>
<td>87.81</td>
<td>0.87</td>
</tr>
<tr>
<td>2000-100-18</td>
<td>86.90</td>
<td>0.86</td>
</tr>
<tr>
<td>500-100-18</td>
<td>83.20</td>
<td>0.82</td>
</tr>
<tr>
<td>2000-50-18</td>
<td>86.62</td>
<td>0.86</td>
</tr>
<tr>
<td>500-50-18</td>
<td>83.88</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Textbox 1. Attention weights of posts. The “text” was paraphrased to anonymize users’ identities.

Users with depression

- User 1
  - Highest weight: I was also dealing with depression and anxiety badly. School was hell.
  - Lowest weight: @user Exam without someone’s supervision is bad.

- User 2
  - Highest weight: I get some rest, take medication, and engage with what I like. These help me and I do not force myself to do things.
  - Lowest weight: Talk about offensive things to physical harm: url.

Users with no depression

- User 1
  - Highest weight: The lady christmas jumper: url.
  - Lowest weight: All the best for your match and hope to see you play.

- User 2
  - Highest weight: He reminds me someone in a football team. He can play many positions and he is our best player.
  - Lowest weight: People believe you when you have evidence.

A recent survey on using social media data to identify users with depression showed that users from the United Kingdom expressed serious concerns about privacy risks and did not see the potential societal benefits outweighing these risks [53]. Thus, if these technologies are to have a meaningful impact on people’s lives, increased importance must be placed on the transparency and trust of the analytics performed. Achieving this trust is, to an extent, helped by the compliance of any research with ethical codes and with the General Data Protection Regulation (GDPR), which helps in raising confidence in data safety and transparent analysis. However, GDPR Article 9: Processing of special categories of personal data specifically mentions that consent is not required if permission relates to personal data that are manifestly made public by the data subject. A core problem is the perception that any data in the public domain are automatically available for research. This is highly controversial from an ethical point of view, as the disruption presented by the wide availability of...
social network data impacts the norms that guide our perception of the usage of our data for research. Ultimately, GDPR is focused on process, not on the objective of the research, which is fundamental to shaping any research consent and the social consensus around it.

This study had some limitations. Collecting control group data is challenging because the samples may contain users with depression who do not publicly express their mental health state on their profiles. Although keyword-based self-declaration is a popular way of asserting depression [11,12], social media users with depression may use more complex ways of communicating their mental health state [54]. There is evidence that social media users post less frequently when they feel low, suggesting that there may be less data available for modeling depression [53].

With regard to technical limitations, this study used additional features from a language analysis tool, which counts words in psychological and word function categories. This may prevent our models from learning word functions directly from sentences. Our future work will use sentence structures extracted from text and train a predictive model with those features [55], which may produce further performance improvements.

The availability of data for model validation is another major concern. Owing to potential ethical issues, there are currently no open data sets to evaluate the performance of predictive models on social network data, making it difficult to compare the model performance. The alternative benchmarking approach used in this study is to replicate well-known study models in the field and apply them to the same data set as the new model being investigated.

Another source of potential bias is the pages that publish tweets about mental health information (eg, mental health charities) and users who report depression experiences of other people (eg, users’ friends, family, or a celebrity). Although we filtered those instances in our study, a significant concern still exists for similar work in the field.

Conclusions
This paper proposes two novel MIL models with and without anaphoric resolution to detect Twitter users with depression. Anaphoric resolution is introduced to address the problem of identifying the subject of a statement made in the post. The classifiers developed comprise a tweet encoder, word attention, tweet classification, user encoder, anaphoric resolution encoder, tweet attention, and user classification layers. Bidirectional long short-term memory layers were used to learn the sequence of words and order of tweets posted on a timeline. Word embedding was applied to transform the textual content into vectors. Additional pronoun features were used to add informative dimensions to our proposed model and highlight posts relevant to the posters themselves. The approach was evaluated against previously published traditional machine learning and deep learning techniques, and the experimental results show that our proposed model produces notably better results. Anaphoric resolution, in particular, improved the performance of our model further and should be considered for inclusion in future studies.

The potential impact of this research lies in its ability to offer social media users exhibiting signs of depression that are suitable for their formal diagnosis. As in other mental health disorders, treatments for depression produce better outcomes and at a lower cost of treatment, the earlier patients get into services. Targeted advertising by mental health charities may be seen as intrusive but is no different than companies advertising any other products to potential consumers based on their web activity.

Early research into public perception of this type of data usage shows that there is public skepticism about this approach. To overcome this animosity toward using social media data for mental health prediction modeling, we believe that future research in this area should focus on explainability and interpretability. We have shown that deep learning MIL models perform well, but they offer no explanation of their decision-making processes [56,57]. Extraction of patterns from the models can provide interpretability, as we demonstrated with tweet weight examples, and systematic sampling should be used to achieve the levels of trust acceptable to users. To gauge how acceptable these techniques are to the public, we intend to work with citizen juries to explore the change in opinion that such explainability can deliver [58].

Acknowledgments
AW is fully funded by a scholarship from the Royal Thai Government to study for a PhD. MAV was supported by Comunidad de Madrid (grants 2016-T1/SOC-1395 and 2020-5A/SOC-19723) and AEI /UE FEDER (grant P12017-85159-P). This work was partially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/P010105/1 (CONSULT: Collaborative Mobile Decision Support for Managing Multiple Morbidities). VC is also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy’s and St Thomas’ National Health Service NHS Foundation Trust and King’s College London, and the Public Health and Multimorbidity Theme of the National Institute for Health Research’s Applied Research Collaboration (ARC) South London. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of the funders.

Conflicts of Interest
None declared.

References
1. Depression. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2021-05-14]


Abbreviations

AIC: Akaike information criterion
API: application programming interface
CNN: convolutional neural network
GDPR: General Data Protection Regulation
GloVe: Global Vectors for Word Representation
GRU: gated recurrent unit
LIWC: linguistic inquiry and word count
MIL: multiple instance learning
MILA-SocNet: multiple instance learning with an anaphoric resolution for social network
MILNET: multiple instance learning network
MIL-SocNet: multiple instance learning for social network
NLP: natural language processing
A Compassion-Focused Ecological Momentary Intervention for Enhancing Resilience in Help-Seeking Youth: Uncontrolled Pilot Study

Christian Rauschenberg1,2, MSc; Benjamin Boecking3, Dipl-Psych, PhD, DClinPsy; Isabell Paetzold1, MSc; Koen Schruers2,4, PhD, MD; Anita Schick1, Dipl-Psych, PhD; Thérèse van Amelsvoort2,4, PhD, MD; Ulrich Reininghaus1,5,6, Dipl-Psych, MSc, PhD

1Department of Public Mental Health, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
2Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands
3Charité, Universitätsmedizin Berlin, Berlin, Germany
4Mondriaan Mental Health Center, Maastricht, Netherlands
5ESRC Centre for Society and Mental Health, King’s College London, London, United Kingdom
6Centre for Epidemiology and Public Health, Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, United Kingdom

Corresponding Author:
Ulrich Reininghaus, Dipl-Psych, MSc, PhD
Department of Public Mental Health
Central Institute of Mental Health
Medical Faculty Mannheim, Heidelberg University
J5
Mannheim, 68159
Germany
Phone: 49 621 1703 1930
Email: ulrich.reininghaus@zi-mannheim.de

Abstract

Background: Digital interventions offer new avenues for low-threshold prevention and treatment in young people. Ecological momentary interventions (EMIs) represent a powerful approach that allows for adaptive, real-time, and real-world delivery of intervention components in daily life by real-time processing of ecological momentary assessment (EMA) data. Compassion-focused interventions (CFIs) may be particularly amenable to translation into an EMI to strengthen emotional resilience and modify putative risk mechanisms, such as stress sensitivity, in the daily lives of young help-seeking individuals.

Objective: This study aims to investigate the feasibility, safety, and initial therapeutic effects of a novel, accessible, transdiagnostic, ecological momentary CFI for improving emotional resilience to stress (EMIcompass).

Methods: In this uncontrolled pilot study, help-seeking youth with psychotic, depressive, or anxiety symptoms were offered the EMIcompass intervention in addition to treatment as usual. The EMIcompass intervention consisted of a 3-week EMI (including enhancing, consolidating, and EMA-informed interactive tasks) administered through a mobile health app and three face-to-face sessions with a trained psychologist intended to provide guidance and training on the CFI exercises presented in the app (ie, training session, follow-up booster session, and review session).

Results: In total, 10 individuals (mean age 20.3 years, SD 3.8; range 14-25) were included in the study. Most (8/10, 80%) participants were satisfied and reported a low burden of app usage. No adverse events were observed. In approximately one-third of all EMAs, individuals scored high on stress, negative affect, or threat anticipation during the intervention period, resulting in real-time, interactive delivery of the CFI intervention components in addition to weekly enhancing and daily consolidating tasks. Although the findings should be interpreted with caution because of the small sample size, reduced stress sensitivity, momentary negative affect, and psychotic experiences, along with increased positive affect, were found at postintervention and the 4-week follow-up. Furthermore, reductions in psychotic, anxiety, and depressive symptoms were found (r=0.30-0.65).
Conclusions: Our findings provide evidence on the feasibility and safety of the EMIconpass intervention for help-seeking youth and lend initial support to beneficial effects on stress sensitivity and mental health outcomes. An exploratory randomized controlled trial is warranted to establish the feasibility and preliminary evidence of its efficacy.

(JMIR Ment Health 2021;8(8):e25650) doi: 10.2196/25650

KEYWORDS

mental health; adolescent psychopathology; digital interventions; mobile health; self-compassion; ecological momentary assessment; mobile phone

Introduction

Background

Most mental disorders first emerge in adolescence and young adulthood (three-fourths by the age of 24 years [1]), with an estimated lifetime prevalence of approximately 50% of any mental disorder in young age groups [1-5]. Furthermore, the Global Burden of Disease study has reported that mental and substance use disorders in children and youth aged 10 to 24 years were the leading cause of overall disease burden in high-income countries [6-8]. Evidence further suggests that most mental disorders are continuous—phenomenologically and temporarily—and, in their early stages, are nonspecific in nature, often evolving into transdiagnostic phenotypes associated with a range of exit psychopathologies [9-16]. Consequently, clinical staging models as an adjunct to formal diagnoses have been introduced [17-19], highlighting the importance of transdiagnostic (indicated) prevention and early intervention [20-24].

Recent transformations in our understanding of the phenomenology, etiology, and early course of mental disorders have contributed to a move toward early detection and prevention [10-13,20,25-31]. Although conventional mental health services offer a range of therapeutic options, it has been widely documented that psychological help remains difficult to access, especially for young individuals in the early stages of mental health problems [21,22,32,33]. Furthermore, tailoring therapeutic options to specific needs and preferences of youth remains a challenge [32-36] and likely contributes to the problem that only a fraction of young people in need of help access any mental health service. Hence, young individuals often experience a long duration of untreated mental health problems, which has been identified as an important marker of poor course and outcome [32].

There is increasing interest in using digital tools to deliver mental health services [37], which may help extend access to and personalization of mental health care [38,39]. This shift has driven the development of novel mobile health (mHealth) interventions for various mental health problems [40-42], of which ecological momentary interventions (EMIs) [23,34,38,39,43], such as the Acceptance and Commitment Therapy in Daily Life [34-36,44], represent a very powerful approach. EMIs allow for adaptive, real-time, and real-world transfer of intervention components in individuals’ daily lives. Thus, EMIs provide a unique opportunity to deliver personalized, precision interventions tailored to what young individuals need in a given moment and context through interactive sampling in real time and the real world. They are based on fine-grained ecological momentary assessment (EMA) data acquired through cutting-edge digital technology [21,23,24,38,39,45,46]. More recently, some authors have started to use the term just-in-time adaptive interventions, which emphasize EMI’s capability of adapting the delivery of intervention components to person and context based on experience sampling or other, for example, sensing data [47,48].

One tangible prevention and early intervention strategy using digital tools is to identify and target transdiagnostic psychological mechanisms in daily life, which have been shown to be involved in the development of mental health problems [23,38]. In recent years, research using EMA—a structured diary technique, also known as experience sampling methodology [43]—has contributed to a better understanding of putative mechanisms likely to impact different stages and increase the intensity of mental health problems in individuals’ daily lives, in real time and outside the research laboratory [21-23,29,43,49,50]. To date, the psychological mechanism most widely studied in daily life is elevated stress sensitivity, characterized by more intense negative affective and psychotic experiences in response to minor stressors and routine daily hassles [22,24,29,43]. Previous studies have suggested that stress sensitivity is elevated in individuals with (1) higher familial or psychometric risk, (2) an ultra–high risk state for psychosis, (3) other early mental health problems, (4) first-episode psychosis, (5) severe and enduring psychosis, and (6) depressive disorders [21,22,24,28,50-58]. In addition, heightened interpersonal sensitivity and threat anticipation have previously been reported to represent further candidate mechanisms in individuals with ultra–high risk state for psychosis, paranoia, and psychotic disorders [24,29,30,59-62] and individuals with depression and anxiety [63-66]. These transdiagnostic mechanisms reflect candidate targets to be modified by EMIs [21,22,24,29].

Compassion-focused interventions (CFIs) are considered an important strand of transdiagnostic interventions for modifying emotion regulation systems [67,68]. CFIs are part of third-wave cognitive behavioral therapy (CBT) and previous meta-analytic evidence on third-wave CBT, including CFIs [69-73], suggest that these types of interventions may yield improvements in mental health outcomes of moderate-to-large effect size. CFIs have been successfully administered to and appraised positively by help-seeking individuals, including individuals with depression, anxiety, and psychosis [74-77]. Furthermore, CFIs have been shown to induce reductions in negative affect and paranoia in moments of high stress in previous research laboratory experimental work [78,79]. In addition, positive imagery, an important component of CFIs, has been found effective in...
reducing various mental health problems, including depression, anxiety, and psychosis [76,80,81] and increasing positive affect, optimism, and behavioral activation [79,82-84]. Thus, CFIIs are particularly well placed to be administered as an EMI to strengthen emotional resilience and modify putative risk mechanisms of poor mental health in young individuals with psychological distress [72,78,85], including stress sensitivity and threat anticipation [21,22]. However, the use of conventional CFIIs under real-world conditions remains limited [86].

As young individuals are digital natives, translating CFI components into an EMI administered through an mHealth app may be a particularly promising approach, offering entirely new avenues for low-threshold prevention and intervention in youth. EMIs are fundamentally translational as they directly build on evidence of underlying momentary mechanisms in daily life and translate these into the development and evaluation of novel digital interventions by targeting these mechanisms in real time and the real world, outside the research lab or clinic [23,39,43]. However, it remains to be established whether evidence on reductions in negative affect and paranoia in moments of high stress—observed in the research laboratory—and effects on other mental health outcomes can indeed be translated to real-world and real-time delivery of EMIs that harness CFI techniques, especially in young help-seeking individuals, where accessible, youth-friendly translation of prevention and early intervention principles reflects a particular challenge.

This Study

The current study aims to establish the clinical feasibility, safety, and initial therapeutic effects of a novel, accessible, transdiagnostic, ecological momentary CFI for improving emotional resilience to stress (EMIcompass) in an uncontrolled phase 1 pilot study in help-seeking youth with psychotic, depressive, or anxiety symptoms. The EMIcompass intervention consisted of a 3-week EMI and three face-to-face sessions with a trained psychologist (ie, training session, follow-up booster session, and review session). Specifically, the intervention offered widely used CFI techniques (eg, compassionate and positive imagery, compassionate writing, and emotion as a wave). To facilitate the interactive, real-time, and real-world translation of the therapeutic content and techniques used in the initial training and booster sessions into individuals’ daily lives, the EMI was administered through an mHealth app on a smartphone. The EMI consisted of (1) enhancing tasks, (2) consolidating tasks, and (3) EMA-informed interactive tasks that aim at an ecological translation of CFI principles and techniques to daily life. Participants were required to complete one enhancing task per week, which allowed them to practice new compassion-focused exercises that were then extended throughout the study period. In addition, they were required to practice the learned CFI components once a day by completing the consolidating tasks. Each time an enhancing task was presented, the intervention components covered by consolidating tasks were expanded. Participants were also offered interactive tasks if they scored high on stress, negative affect, or threat anticipation in daily EMA. The face-to-face sessions were designed to provide guidance and training on the CFI exercises and how to use the app, background information on the strategies presented, and discussions of open questions and challenges participants encountered while using the app.

The primary objective of this study is to (1) assess the clinical feasibility of delivering the EMIcompass intervention to help-seeking youth based on successful recruitment, assessment of outcomes, compliance, satisfaction, and acceptability and safety by carefully documenting any serious adverse events throughout the study period. The secondary objectives were to examine (2) initial therapeutic effects of EMIcompass on reducing stress sensitivity, negative affect, and psychotic experiences, and increasing positive affect in daily life at the end of the 3-week intervention period (postintervention), and after a 4-week follow-up period (follow-up), along with (3) the initial therapeutic effects of EMIcompass on reducing threat anticipation, psychotic, depressive, and anxiety symptoms as well as general psychopathology.

Methods

Study Design

In an uncontrolled phase 1 pilot study, help-seeking individuals with psychotic, depressive, or anxiety symptoms aged between 14 and 25 years were referred to secondary mental health services in the Netherlands (ie, Mondriaan Mental Health Trust and Virenze Mental Health Care) and received the EMIcompass intervention in addition to treatment as usual. Data were collected before the intervention (baseline), at the end of the 3-week intervention period (postintervention), and after a 4-week follow-up period (follow-up). Close attention was paid to establishing the clinical feasibility (eg, pragmatic inclusion and exclusion criteria based on routine assessments) and safety (ie, documentation of any serious adverse events) of this study. Our recruitment strategy drew on our previous and ongoing work with youth [22,24,29,34-36,44] and guidance for pragmatic randomized controlled trials (RCTs) [87] and hence was geared to reflect the heterogeneity of the population commonly encountered in routine care.

Sample

We recruited young individuals with psychotic, depressive, and/or anxiety symptoms who sought help from two secondary mental services (ie, Mondriaan Mental Health Trust and Virenze Mental Health Care). The inclusion and exclusion criteria were equivalent in principle across the two services but were purposefully selected to be pragmatic and hence based on routine assessments for screening, diagnosis, formulation, and outcome measurement, which differed between the two services (Textbox 1). This approach was adopted to ensure that the aim of establishing feasibility reflected the population actually encountered in clinical practice (rather than imposed by researchers) while keeping the assessment burden at a minimum. The study was approved by the Ethics Review Committee of Mondriaan Mental Health Center and the Ethics Review Committee of Psychology and Neuroscience, Maastricht University. A flowchart of the study is shown in Figure 1.

The prodromal questionnaire (PQ) [88,89], which has been reported to be a very good screening measure in routine mental health services [89,90], was used to screen for psychotic...
symptoms. In addition, the Brief Symptom Inventory (BSI) [91,92] was used to screen for anxiety, depressive, and psychotic symptoms, and the Symptom Questionnaire-48 [93] was used in addition to the PQ to screen for anxiety and depressive symptoms.

Textbox 1. Inclusion and exclusion criteria by participating mental health services.

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Mondriaan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aged between 18 and 25 years</td>
</tr>
<tr>
<td></td>
<td>Prodromal questionnaire score of 6 or above</td>
</tr>
<tr>
<td></td>
<td>Symptom questionnaire-48 score of 9 or above on the social phobia subscale, or score of 8 or above on the depression subscale, or score of 11 or above on the anxiety subscale</td>
</tr>
<tr>
<td></td>
<td>Willingness to participate in the compassion-focused ecological momentary intervention</td>
</tr>
<tr>
<td></td>
<td>Ability to give written informed consent independently, without help from others</td>
</tr>
<tr>
<td>Virenze</td>
<td>Aged between 14 and 25 years</td>
</tr>
<tr>
<td></td>
<td>Prodromal questionnaire score of 6 or above</td>
</tr>
<tr>
<td></td>
<td>Brief Symptom Inventory t score of 63 or above</td>
</tr>
<tr>
<td></td>
<td>Willingness to participate in the compassion-focused ecological momentary intervention</td>
</tr>
<tr>
<td></td>
<td>Ability to give written informed consent independently, without help from others</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion Criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Insufficient command of Dutch, primary clinical diagnosis of alcohol or substance dependency, severe endocrine, cardiovascular, or organic brain disease</td>
</tr>
</tbody>
</table>

The EMIcompass Intervention

Development of the Manual

The intervention was structured and manualized to ensure consistent delivery. The manual was based on widely used CFI techniques (eg, compassionate and positive imagery, compassionate writing, and emotion as a wave) and developed following a process of reviewing existing manuals and the extant CFI literature [67,68,73,74,78,80] through the team’s clinical experience of working with these approaches with clients and through consultation with local experts in CFI and the wider research team. The intervention was designed based on the principles of EMIs [23,34-36,39,43,44].

EMIcompass Intervention and Treatment as Usual

In this study, participants were offered the EMIcompass intervention in addition to treatment as usual, which included all the treatment they received before the start of the study (ie, good standard care delivered according to local and national guidelines by their general practitioner, psychiatrist, and other health care professionals), including CBT, third-wave CBT, dialectical behavior therapy, and other psychological interventions. The EMIcompass intervention consisted of three face-to-face sessions (one training session, one follow-up booster session, and one review session) given by a trained psychologist, who was supervised by an expert clinical psychologist in compassion-focused therapy, and a 3-week EMI administered through an mHealth app on a smartphone (PsyMate; Psymate BV). In addition, participants were offered on-demand email and/or phone contact during the intervention period.

At the beginning of the 3-week intervention period, an initial face-to-face training session was offered to participants. This session was fully manualized based on previous research that used CFIs [67,68,74,78,94]. The goal of the first session was to train individuals to cope with negative emotions by applying a personal, compassionate image that conveys compassion, care, and warmth to them based on the description of Gilbert [68], as applied by Lincoln et al [78]. This was followed by inducing negative emotions using in-sensu exposure to a personally relevant social situation that participants remember having experienced as distressing. This method has been safely applied to individuals with mental health problems [74,78] without any adverse consequences or health-related risks. Following the induction of negative emotions, participants were asked to practice a 5-minute application of the compassionate image they were trained in at the beginning of the session [67,68]. Training the use of compassionate imagery was repeated and extended to imagery involving a compassionate self [68] and emotion as a wave [94] in the following booster session 2 weeks after the initial training session. In the review session at the end of the 3-week intervention period, the smartphone was returned, and progress
and satisfaction with and acceptability of the intervention were reviewed and assessed. To allow for interactive, real-time, and real-world translation of the therapeutic content and techniques of initial and booster sessions into individuals’ daily lives, participants were offered a 3-week EMI delivered through an mHealth app. During the 3-week intervention period, the smartphone prompted a signaling sound from the smartphone seven times per day on 6 consecutive days per week to reduce the burden associated with app usage. At each beep, participants were asked to complete a brief EMA on Momentary stress, positive and negative affect, and threat anticipation in daily life (see the section on EMA measures used). The EMA was scheduled at random within set blocks of time. The EMI consisted of 3 different types of tasks (Table 1): participants were asked to complete one *enhancing task* per week, allowing them to practice new compassion-focused exercises, which were subsequently extended during the study period (eg, discovering their own compassionate self and experiencing emotions as a wave). In addition, they were asked to practice the learned CFI components once a day by completing the *consolidating tasks* at a predefined time. The components covered by consolidating tasks were extended each time an enhancing task was presented. Furthermore, *interactive tasks* were offered if participants scored highly on stress, negative affect, or threat anticipation in the EMA (ie, scores higher than 4 on a 7-point Likert scale). As an essential element of compassion-focused therapy is the use of compassionate imagery in moments of high stress, negative affect, or threat anticipation, these interactive tasks are thought to reflect a core active component of the 3-week compassion-focused EMI.

### Table 1. Components of the EMICompass intervention.

<table>
<thead>
<tr>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Compassion-focused training sessions</strong></td>
<td>Training session (compassionate image)</td>
<td>Booster session (day 11-15; compassionate self-training, “emotion as a wave”)</td>
</tr>
<tr>
<td><strong>Enhancing tasks</strong></td>
<td>Task 1 (day 3 or 4): compassionate self-validation</td>
<td>Task 2 (day 9 or 10): “emotion as a wave”</td>
</tr>
<tr>
<td><strong>Consolidating tasks</strong></td>
<td>Compassionate self-validation (from day 5, following enhancing EMI task 1)</td>
<td>Companionship self-validation “Emotion as a wave” (from day 11, following enhancing EMI task 2)</td>
</tr>
<tr>
<td><strong>Interactive tasks</strong></td>
<td>Compassionate image Compassionate self-validation (from day 5, following enhancing EMI task 1)</td>
<td>Compassionate image Compassionate self-validation “Emotion as a wave” (from day 11, following enhancing EMI task 2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aEMI: ecological momentary intervention.*

### Measures

**Sociodemographic Characteristics**

A sociodemographic schedule was used to assess age, gender, occupation, and level of education.

**Clinical Feasibility and Safety**

Feasibility was assessed based on successful recruitment, assessment of outcomes, compliance with the manual, satisfaction, and acceptability. For some of the feasibility domains, a debriefing scale was used. The reasons participants declined to participate in the study were carefully recorded, and the completeness of outcomes at each time point was documented. Acceptability was assessed in the review session of the EMICompass intervention together with the trained psychologist by asking participants to complete a feedback form about the EMI tasks and sessions and rate the extent to which they felt they benefited from and were satisfied with the intervention [74,78]. In addition, the trained psychologist asked participants in the review session to report whether they perceived the face-to-face sessions, compassion-focused exercises, and EMI tasks as helpful. App usability was assessed by asking participants to rate the readability of the text shown on the screen, any difficulties in operating the app or technical problems, the clarity of provided instructions, and whether the app was perceived as burdensome. All items were rated on a 7-point Likert scale ranging from *not at all* (rating of 1) to *moderate* (rating of 4) and *very* (rating of 7), which were subsequently grouped into three categories of *not* (rating of 3 or lower), *moderate* (rating of 4 or 5), and *very* (rating of 6 or 7) for the sake of interpretability of findings (given small numbers in each cell). Safety was assessed by carefully documenting any serious adverse events throughout the entire study period and the potential negative effects of app usage on mental health in participants.

**Stress Sensitivity, Negative and Positive Affect, and Psychotic Experiences in Daily Life**

EMA was used to assess stress sensitivity, negative and positive affect, psychotic experiences, and threat anticipation in daily life. For this, the same app was used as for the EMICompass intervention (PsyMate), and assessments were completed at baseline, postintervention, and 4-week follow-up for 6 consecutive days, following the protocol from previous EMA.
Stress was operationalized as minor disturbances and distinctive unpleasant events, activities, and social situations that occur in the flow of daily life. Event-related stress was measured with an item asking participants to rate the most important event that had happened since the last beep on a 7-point Likert scale ranging from very unpleasant (rating of −3) to very pleasant (rating of 3) [54]. The item was recoded, such as higher ratings indicated higher levels of stress (with ratings of −3 coded as 7 and ratings of 3 coded as 1). Activity-related stress was measured by asking participants first to specify their current activity (eg, resting and watching TV), which was followed by asking them to rate the pleasantness of this activity on a 7-point Likert scale (1=very unpleasant; 7=very pleasant). Social stress was measured by asking participants to specify categorically with whom they were spending time (eg, nobody, partner, or family) and appraise the current social context using the items “I find being with these people pleasant” (reversed), “I feel accepted” (reversed), and “I feel excluded (if with someone)” or “I find it pleasant to be alone” (reversed) and “I would prefer to have company” (if alone) ranging from not at all (rating of 1) to very much (rating of 7). The good concurrent validity of these EMA stress measures has been reported [54,55]. Furthermore, a composite stress score was calculated using the mean score of all seven stress items [21,95]. Negative affect was assessed using five items asking participants to rate the extent to which they felt anxious, down, insecure, uncomfortable, and guilty at each entry point [54]. Positive affect was assessed by asking participants to rate the extent to which they felt cheerful and relaxed, all rated on a 7-point Likert scale ranging from not at all (rating of 1) to very much (rating of 7) [54,55,96]. Psychotic experiences were assessed using seven items (“I see things that aren’t really there,” “I hear things that aren’t really there,” “I feel suspicious/paranoid,” “I feel unreal,” “My thoughts are influenced by other,” “I can’t get these thoughts out of my head,” and “I feel like I am losing control”) rated on a 7-point Likert scale ranging from 1 (not at all) to 7 (very much) [55,96]. Threat anticipation was assessed by asking participants to think of what might happen in the next few hours and rate the item “I think that something unpleasant will happen” on a 7-point Likert scale (ranging from 1=not at all to 7=very much) [24,29]. Negative and positive affect, psychotic experiences, and threat anticipation scores were assessed by computing the mean scores. In line with earlier studies [22,24,29,46,49], items on stress, negative affect, and psychotic experiences were used as a proxy for individuals’ stress sensitivity in daily life by modeling the association between stress and (1) negative affect and (2) psychotic experiences. Thus, we conceptualized stress sensitivity in daily life as individuals’ affective and psychotic reactivity to minor daily stressors.

Psychotic, Depressive, and Anxiety Symptoms and General Psychopathology

We used non-EMA outcome measures to assess psychotic, depressive, and anxiety symptoms and general psychopathology. First, the BSI was used to assess depressive and anxiety symptoms (based on the respective BSI subscales) and general psychopathology by computing the Global Severity Index (based on 53 BSI items). Participants rated each item on a 5-point scale ranging from 0 (not at all) to 4 (extremely) [91,92]. Second, the Green et al, Paranoid Thoughts Scale, a reliable and valid scale, was used to assess psychosis [97]. The Green et al, Paranoid Thoughts Scale was modified to ask participants about paranoid ideation during the past week rather than the past month, given that the intervention period was only 3 weeks. A total score was computed using all 32 items (both with a 5-point scale: 1=not at all, 3=somewhat, and 5=totally). Third, the threat anticipation measure [98] was used to measure threat anticipation by asking participants to estimate the future likelihood of a list of threatening, neutral, and positive events happening to themselves and other people [62,98,99]. Items for threatening and neutral events were used to compute the total scores. Each event was rated separately for the likelihood that it will happen to oneself and another person on a 7-point scale (1=not at all; 7=very likely), resulting in four total sum scores (ie, threat anticipation-self, threat anticipation-other, neutral anticipation-self, and neutral anticipation-other), where higher scores indicate higher probability estimates. Finally, the PQ [88,89] was used to assess the presence of prodromal and attenuated psychotic symptoms (ie, positive symptoms, disorganized symptoms, negative symptoms, and general symptoms). This measure consists of 16 items that assess the presence of psychotic symptoms (0=not present and 1=present), which were used to compute a total score (range 0-16). Good psychometric properties have been reported for these measures [88,97,98,100,101].

Statistical Analysis

STATA 15.1 (StataCorp) was used to analyze the data. First, descriptive statistics were used, and CIs were constructed, as appropriate, to summarize the findings on feasibility and safety. Second, as EMA data have a multilevel structure, such that multiple observations (level 1) are nested within subjects (level 2), linear mixed models were used to control for within-subject clustering of multiple observations using the mixed command in STATA. Thus, to examine the effects of the EMIcompass intervention on reducing stress sensitivity, EMA stress variables and time points were included as independent variables and negative affect and psychotic experiences as the outcome variable in linear mixed models, which were fitted separately for each outcome variable. We then added two-way interaction terms for stress×time and used likelihood ratio tests (lrtest command) to evaluate improvement in model fit and the lincom command to compute linear combinations of coefficients to test our hypotheses on whether stress sensitivity was reduced at postintervention and the 4-week follow-up. We standardized continuous ESM (experience sampling method) variables (mean 0, SD 1) to interpret significant interaction terms. Family-wise error-corrected P values were computed to control for multiple testing by multiplying the unadjusted P values of the two-way interaction effects by the total number of tests (N=4) for each outcome. Third, to examine the effects of the EMIcompass intervention on other EMA outcome measures, time points were included as independent variables and negative affect, positive affect, psychotic experiences, and threat anticipation as the outcome variable in separate linear mixed models. All models were controlled for potential confounders (ie, age, gender, and level of education). Finally, we used Wilcoxon signed-rank tests.
to examine the effects of EMIcompass on non-EMA outcome measures of threat anticipation, psychotic, depressive, and anxiety symptoms and general psychopathology at postintervention and 4-week follow-up. The resulting z scores were used to calculate the effect sizes displayed in $r$ as described by Rosenthal and DiMatteo [102].

**Results**

**Sociodemographic and Clinical Characteristics**

A flowchart of the study is shown in Figure 1 and basic sample characteristics in Table 2. In total, 30 potential participants aged between 14 and 25 years were referred to the study by clinicians from the two participating mental health services. Of these, 16 provided written informed consent and were eligible, of whom 11 completed the baseline assessment and were included in the EMIcompass intervention. A participant was lost during the 3-week intervention period, whereas 10 participants (mean age 20.3 years, SD 3.8; range 14-24) completed the EMIcompass intervention and both postintervention and 4-week follow-up assessments. Most participants were women (7/10, 70%) and were currently at school/university (6/10, 60%). Half of the participants had a clinical diagnosis of major depressive disorder (5/10, 50%) and met the criteria for a comorbid mental health condition. Most participants were of White Dutch ethnicity, and some reported having used cannabis during the previous 12 months (3/10, 30%).
Table 2. Basic sample characteristics of service users (N=10).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>20.3 (3.8; 14-25)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Male</td>
<td>3 (30)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>White Dutch</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Missing value</td>
<td>3 (30)</td>
</tr>
<tr>
<td><strong>Level of education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Further</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Higher</td>
<td>4 (40)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>School or education</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Employed (full- or part-time)</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Unstructured activities</td>
<td>1 (10)</td>
</tr>
<tr>
<td><strong>Cannabis use</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Lifetime</td>
<td>4 (40)</td>
</tr>
<tr>
<td><strong>DSM-IV diagnosis, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>5 (50)</td>
</tr>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Reactive attachment disorder</td>
<td>2 (20)</td>
</tr>
<tr>
<td>None</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Comorbid condition^d</td>
<td>5 (50)</td>
</tr>
</tbody>
</table>

^aCategories defined as school (elementary school), further (voorbereidend middelbaar beroepsonderwijs [VMBO]), hoger algemeen voortgezet onderwijs [HAVO], and voorbereidend wetenschappelijk onderwijs [VWO], and higher (hoger beroepsonderwijs [HBO], and wetenschappelijk onderwijs [WO]) of the Dutch educational system.

^bOn the basis of Composite International Diagnostic Interview section of Illegal Substance Use and defined as having used cannabis more than five times on its own initiative during the previous 12 months or lifetime.

^cDSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

^dConsisting of the following diagnostic categories: panic disorder, attention-deficit/hyperactivity disorder, intermittent explosive disorder, borderline personality disorder, and parent-child relational problem.

Clinical Feasibility and Safety

The clinical feasibility and safety findings are shown in Table 3. Almost all individuals (9/10, 90%) reported that participating in the study did not interfere with their daily activities. Most individuals reported being very (40%-50%) or moderately satisfied (40%-50%) with tasks delivered through the EMCompass app and moderately (20%-30%) or very (60%) satisfied across face-to-face sessions. Most participants were also very (5/10, 50%) or moderately (2/10, 20%) successful in imagining a compassionate image. Some individuals reported that the intervention positively influenced social contacts (3/10, 30%; ratings of moderate and very combined) and levels of activity (4/10, 40%). All individuals were very satisfied with the face-to-face contact sessions and felt trained psychologists understood them. Although all participants reported that they were able to follow the instructions shown on the screen, observer ratings by trained psychologists, who also delivered the face-to-face sessions, indicated that some individuals might have had problems with this (1/10, 10% in session 1 and 2/20, 20% in session 3). Findings on app usability were satisfactory, and the burden associated with app usage was perceived to be low or very low across all time points (70%-90%), although some individuals (3/10, 30%) found the number of signals per day to be moderately burdensome. In addition, some individuals perceived the items used in the PsyMate app as difficult or
unclear (2/10, 20%). No severe adverse events were observed during the study period.

In-app usage data during the intervention period suggest high completion rates of EMA assessments. Specifically, the EMIcompass app triggered 1260 signals asking participants to complete brief EMA assessments (126 for each person). Of these 1260 signals, individuals reacted to 467 (37.06%), although high variability between individuals was found (range 214/1260, 16.9% to 844/1260, 66.9%). Individuals scored high on stress, negative affect, or threat anticipation in 32.1% (150/467) of EMA assessments, resulting in real-time delivery of CFI intervention components in approximately 1 out of 3 of all completed EMA assessments. When considering the assessment of outcomes at baseline, postintervention, and follow-up, we found satisfactory compliance rates (no missing data for outcome measures filled in person and at least 18/60, 30% of all EMA assessments). Thus, when combining self-reports and in-app usage data, assessing outcomes and compliance with the manual was considered satisfactory. Furthermore, the conversion rate of recruitment was 3:1 (ie, from identified to included individuals; Figure 1), which is in line with previous research and considered successful recruitment.
Table 3. Findings on safety, feasibility, and app usability of the EMIcompass intervention.

<table>
<thead>
<tr>
<th>Safety and feasibility, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interference of study participation with daily activities</td>
</tr>
<tr>
<td>Satisfaction with face-to-face sessions</td>
</tr>
<tr>
<td>Session 1: compassionate image; inducing negative emotions</td>
</tr>
<tr>
<td>Session 2: compassionate self; emotion as a wave</td>
</tr>
<tr>
<td>Session 3: review session</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Satisfaction with tasks, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1: compassionate self-validation</td>
</tr>
<tr>
<td>Task 2: emotion as a wave</td>
</tr>
<tr>
<td>Task 3: self-compassionate writing</td>
</tr>
<tr>
<td>Self-reported success in making a compassionate image</td>
</tr>
<tr>
<td>Taking part in the study positively affected activities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taking part in the study affected social contacts, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positively</td>
</tr>
<tr>
<td>Negatively</td>
</tr>
<tr>
<td>Satisfaction with contact with trained psychologist</td>
</tr>
<tr>
<td>Participant felt understood by trained psychologist</td>
</tr>
<tr>
<td>Self-reported level of understanding of instructions provided by trained psychologist</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observer-rating by trained psychologists, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compliance in session 1</td>
</tr>
<tr>
<td>Compliance in session 2</td>
</tr>
<tr>
<td>Compliance in session 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMIcompass app usability, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Readability of text on screen</td>
</tr>
<tr>
<td>Difficulties in operating the app</td>
</tr>
<tr>
<td>Clarity of instructions given on screen</td>
</tr>
<tr>
<td>Difficulties understanding used items</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMIcompass app perceived as burdensome, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In terms of the number of signals per day</td>
</tr>
<tr>
<td>In terms of the number of items asked per signal</td>
</tr>
<tr>
<td>In terms of the signal sound</td>
</tr>
<tr>
<td>Technical problems</td>
</tr>
</tbody>
</table>

---

*Items were rated on a 7-point Likert scale ranging from not at all (rating of 1) to moderate (rating of 4) and very (rating of 7). Trained psychologists noted the answers. The answers were grouped into three categories of not (rating of 3 or lower), moderate (rating of 4 or 5), and very (rating of 6 or 7) for the sake of interpretability (given small numbers in each cell).*

*Missing value for 1 participant.*
**Initial Therapeutic Effects**

**Stress Sensitivity, Negative and Positive Affect, and Psychotic Experiences in Daily Life**

The findings on the initial therapeutic effects of the EMIcompass intervention on stress sensitivity are provided in Table 4. We found preliminary evidence that participants experienced less intense negative affect in response to event-related and activity-related stress at postintervention and in response to overall, event-related, activity-related, and social stress at follow-up than at baseline, as indicated by statistically significant two-way interaction effects for stress×time point. Furthermore, participants reported less intense psychotic experiences in response to minor stressors in daily life (ie, overall and specific types of stressors) at postintervention and follow-up than at baseline.

**Table 4.** Initial therapeutic effects of EMIcompass on stress sensitivity in daily life.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Postintervention versus baseline</th>
<th>Follow-up versus baseline</th>
<th>Follow-up versus postintervention</th>
<th>Likelihood ratio test for interaction(^a)</th>
<th>Chi-square (df)</th>
<th>PFWE(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negative affect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>−0.12 (−0.27 to 0.03)</td>
<td>.11</td>
<td>−0.51 (−0.63 to −0.40)</td>
<td>&lt;.001</td>
<td>−0.39 (−0.55 to −0.23)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Event-related</td>
<td>−0.41 (−0.56 to −0.25)</td>
<td>&lt;.001</td>
<td>−0.39 (−0.51 to −0.27)</td>
<td>&lt;.001</td>
<td>0.02 (−0.14 to 0.18)</td>
<td>.83</td>
</tr>
<tr>
<td>Activity-related</td>
<td>−0.25 (−0.40 to −0.09)</td>
<td>.002</td>
<td>−0.35 (−0.47 to −0.23)</td>
<td>&lt;.001</td>
<td>−0.10 (−0.27 to 0.06)</td>
<td>.22</td>
</tr>
<tr>
<td>Social</td>
<td>0.05 (−0.10 to 0.20)</td>
<td>.50</td>
<td>−0.41 (−0.53 to −0.28)</td>
<td>&lt;.001</td>
<td>−0.46 (−0.62 to −0.29)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Psychotic experiences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>−0.15 (−0.25 to −0.04)</td>
<td>.005</td>
<td>−0.28 (−0.36 to −0.20)</td>
<td>&lt;.001</td>
<td>−0.14 (−0.25 to −0.03)</td>
<td>48.7 (2)</td>
</tr>
<tr>
<td>Event-related</td>
<td>−0.29 (−0.39 to −0.19)</td>
<td>&lt;.001</td>
<td>−0.19 (−0.27 to −0.11)</td>
<td>&lt;.001</td>
<td>0.10 (−0.01 to 0.20)</td>
<td>.08</td>
</tr>
<tr>
<td>Activity-related</td>
<td>−0.25 (−0.35 to −0.14)</td>
<td>&lt;.001</td>
<td>−0.20 (−0.28 to −0.12)</td>
<td>&lt;.001</td>
<td>0.05 (−0.06 to 0.16)</td>
<td>.40</td>
</tr>
<tr>
<td>Social</td>
<td>−0.01 (−0.11 to 0.09)</td>
<td>.86</td>
<td>−0.24 (−0.32 to −0.16)</td>
<td>&lt;.001</td>
<td>−0.23 (−0.34 to −0.12)</td>
<td>36.3 (2)</td>
</tr>
</tbody>
</table>

\(^a\)Likelihood ratio test for stress×time interaction after inclusion in the following model: (for y\(_{ij}\) negative affect, psychotic experiences or positive affect as outcome variable): y\(_{ij}\) = β\(_0\) + β\(_1\)(STRESS\(_j\)) + β\(_2\)(TIME\(_j\)) + β\(_3\)(STRESS\(_j\)×TIME\(_j\)) + ε\(_{ij}\).

\(^b\)Adjusted β: standardized regression coefficients (continuous independent variables were standardized [mean 0, SD 1] for interpreting interaction terms).

\(^c\)PFWE: family-wise error-corrected P values were computed by multiplying the unadjusted P value by the total number of tests for each outcome (N=4) to adjust significance levels of likelihood ratio tests for two-way interactions.

Furthermore, Table 5 shows the findings of the initial effects of EMIcompass on momentary negative affect, psychotic experiences, and positive affect. There was preliminary evidence that participants experienced less intense negative affect and psychotic experiences and more intense positive affect in daily life at postintervention and the 4-week follow-up than at baseline. There was also evidence that individuals anticipated fewer threatening events in their daily lives at postintervention and the 4-week follow-up than at baseline.
Table 5. Initial therapeutic effects of EMIcompass on individuals’ momentary stress, negative affect, psychotic experiences, positive affect, and threat anticipation.

<table>
<thead>
<tr>
<th></th>
<th>Baseline, mean (SD)</th>
<th>Postintervention, mean (SD)</th>
<th>Follow-up, mean (SD)</th>
<th>Postintervention versus baseline</th>
<th>Follow-up versus baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta$ (95% CI)</td>
<td>$P$ value</td>
</tr>
<tr>
<td>Positive affect</td>
<td>3.9 (1.8)</td>
<td>4.5 (1.5)</td>
<td>4.3 (1.6)</td>
<td>0.39 (0.16 to 0.62)</td>
<td>.001</td>
</tr>
<tr>
<td>Negative affect</td>
<td>2.2 (1.3)</td>
<td>1.8 (1.1)</td>
<td>1.4 (0.7)</td>
<td>−0.44 (−0.59 to −0.30)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Psychotic experiences</td>
<td>1.7 (0.8)</td>
<td>1.4 (0.9)</td>
<td>1.3 (0.6)</td>
<td>−0.25 (−0.34 to −0.16)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Threat anticipation</td>
<td>2.7 (1.9)</td>
<td>2.2 (1.3)</td>
<td>1.6 (1.1)</td>
<td>−0.61 (−0.83 to −0.39)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Psychotic, Depressive, Anxiety Symptoms, and General Psychopathology

The findings on the initial therapeutic effects of EMIcompass on non-EMA outcome measures are presented in Table 6. Overall, reductions in threat anticipation, psychotic, depressive, and anxiety symptoms and general psychopathology (as indexed by the Global Severity Index) of moderate-to-large effect sizes were found at the end of the 3-week intervention period (postintervention) and after a 4-week follow-up period ($r=0.30$-$0.65$). There was initial evidence, despite the small sample size and, hence, limited statistical power, that these reductions were beyond what would be expected by chance alone for psychotic symptoms at postintervention and 4-week follow-up and, at trend level, for anxiety symptoms (postintervention, 4-week follow-up) and anticipation of a positive future self (4-week follow-up). The intervention effects on depressive symptoms and general psychopathology were also of medium-to-large effect size but fell short of statistical significance. Reductions in threat anticipation (self or other) were only of small-to-moderate effect size and did not reach conventional levels of statistical significance.
Table 6. Initial therapeutic effects of EMIcompass intervention on psychotic, depressive, and anxiety symptoms, general psychopathology, and threat anticipation.

<table>
<thead>
<tr>
<th></th>
<th>Scores, median (range)</th>
<th>Paired Wilcoxon signed-rank test (N=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Postintervention</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Severity Index</td>
<td>81 (22-146)</td>
<td>68.5 (5-158)</td>
</tr>
<tr>
<td>Depression</td>
<td>13.5 (1-23)</td>
<td>12 (0-23)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>11.5 (4-16)</td>
<td>9.5 (0-17)</td>
</tr>
<tr>
<td>BSIb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>41 (32-73)</td>
<td>46.5 (32-83)</td>
</tr>
<tr>
<td>Prodromal questionaire</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>5 (1-10)</td>
<td>5 (0-9)</td>
</tr>
<tr>
<td>TAMf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Future self (positive)</td>
<td>26.5 (17-37)</td>
<td>27 (16-37)</td>
</tr>
<tr>
<td>Future self (threatening)</td>
<td>15.5 (11-25)</td>
<td>16.5 (7-24)</td>
</tr>
<tr>
<td>Future others (positive)</td>
<td>31.5 (19-45)</td>
<td>31 (27-42)</td>
</tr>
<tr>
<td>Future others (threatening)</td>
<td>15.5 (7-37)</td>
<td>14 (8-36)</td>
</tr>
</tbody>
</table>

aEffect size estimates are based on r described by Rosenthal and DiMatteo [102] using the following formula: r=Z/√number of pairs.
bBSI: Brief Symptom Inventory.
cP<.10.
dGPTS: Green et al, Paranoid Thoughts Scale.
eP<.05.
fTAM: threat anticipation measure.

Discussion

Principal Findings

The findings of this uncontrolled phase 1 pilot study suggest initial results on the feasibility, safety, and preliminary therapeutic effects of a compassion-focused ecological momentary transdiagnostic intervention designed to improve emotional resilience to stress (EMIcompass) in help-seeking youth with psychotic, depressive, or anxiety symptoms. First, individuals were satisfied with face-to-face and app-based intervention components, interference with daily activities was low, and observer-rated compliance with the treatment was high. The indicators of app usability were satisfactory. In addition, no adverse effects were observed. Second, there was preliminary evidence of decreased stress sensitivity, negative affect, and psychotic experiences and increased positive affect in daily life at the end of the 3-week intervention period (postintervention) and after a 4-week follow-up period (follow-up) as compared with baseline. Third, there was initial evidence, despite the small sample size and limited statistical power, of reductions in threat anticipation, psychotic, anxiety, and depressive symptoms of medium-to-large effect size (r=0.30-0.65). Overall, this reflects promising preliminary evidence of clinical feasibility and safety of the EMIcompass intervention in help-seeking youth and some evidence on initial therapeutic effects. However, findings on clinical outcomes should be interpreted with caution, considering the small sample size of this pilot study.

Strengths and Limitations

The strength of this study is that the principles of CFIs were, for the first time, translated into an EMI administered through an mHealth app as a new avenue for real-world and real-time prevention and intervention in youth. Furthermore, EMIcompass transforms evidence on putative underlying mechanisms into an intervention that directly targets these mechanisms in daily life and hence is translational. However, there are a number of limitations that must be considered when interpreting our findings. First, in line with state-of-the-art guidance on developing and evaluating complex interventions [103], mHealth interventions in particular [104], the sample size (N=10) of this pilot study was selected to be small. Thus, the primary focus of this study was to investigate feasibility and safety and estimate
the effect size of initial therapeutic effects rather than statistical significance to provide the basis for a feasibility RCT [105]. Nonetheless, while considering the low statistical power and limitations associated with a small sample size, we found preliminary evidence (in terms of statistical significance) on the effects of the EMIconpass intervention on stress sensitivity. These are promising findings, as stress sensitivity is the primary target of this emotion regulation–focused intervention. Second, data on feasibility and acceptability were assessed together with or by a trained psychologist and not an independent person. Thus, we cannot rule out biases and underreporting of unhelpful experiences. Third, we used a modified version of an established debriefing scale already used for a decade in EMA studies and, more recently, in other EMIs [34,35] to assess satisfaction, engagement, and other domains of feasibility. However, the convergent validity of this measure with other established measures (eg, Mobile App Rating Scale) and other psychometric properties remain to be established. Fourth, because of the absence of a waiting list or active control group, we cannot rule out that there may be no additive therapeutic effects of the EMIconpass intervention to the therapeutic effects of the face-to-face sessions with the trained psychologists or other therapeutic interventions participants received during the intervention period in the form of treatment as usual. However, the primary aim of this pragmatic phase 1 pilot study was to provide the basis for a feasibility RCT by investigating feasibility and safety, generating initial effect sizes. Further examination of the efficacy of EMIconpass intervention is urgently warranted. Fifth, most participants were women, and half of the participants had depression, which may limit the generalizability of findings, as selection bias may have operated on our sampling procedure. Sixth, after written informed consent was obtained and baseline assessments were completed, 5 individuals decided not to participate in the study. The reasons for exclusion were not assessed, which limited our findings on feasibility. Finally, the complex nature of the investigated constructs, sample size, and study design exclude any form of causal inference.

**Ideas for Future Work**

The EMIconpass intervention aimed to augment current treatment options for young individuals seeking help for mental health problems. Most individuals reported being satisfied with the intervention. Although the small sample size has to be considered when interpreting findings, the preliminary therapeutic effects on candidate psychological mechanisms, including stress sensitivity and other psychopathological outcomes, were promising. Importantly, no adverse effects have been reported, and participating in the study did not hinder individuals in their daily activities. Thus, overall, findings on feasibility, safety, and initial therapeutic effects may be considered encouraging.

This is one of the first studies to develop and pilot an EMI that incorporates an adaptive and context-dependent delivery scheme of intervention components in youth with mental health problems. The interactive tasks were triggered in approximately 1 out of 3 of all EMA assessments when individuals experienced elevated levels of negative affect (eg, feeling anxious, insecure, down; ie, scores higher than 4 on a 7-point Likert scale) or momentary stress. Thus, real-time data processing was successfully applied based on EMA data to determine the delivery of CFI components. This may represent not only an important step toward ecologically more valid and accessible psychological interventions in youth but also a more personalized and contextualized clinical and preventive approach. In other words, the principles of EMIs allow not only to translate intervention components targeting candidate momentary mechanisms and contexts to individuals’ daily lives but also take a personalized, adaptive approach informed by fine-grained real-time EMA data to produce sustainable change in the real world. Although a feasibility RCT is needed as a significant next step to investigate the efficacy of the intervention and feasibility as a basis for a confirmatory RCT [23,34], this pilot study of this novel EMI reflects an important stepping stone toward more personalized and accessible youth mental health care. Furthermore, in-app data analytics revealed high variability in compliance among individuals. This suggests that for some individuals, the number of signals per day was too high (ie, seven times per day on 6 consecutive days per week).

These findings hint toward potential avenues for the improvement of the EMIconpass intervention to be iteratively incorporated. First, future versions of the EMIconpass intervention may offer adaptive intervention trajectories that vary in the type of exercise depending on individual needs and preferences. Importantly, in doing so, potentially influencing factors (eg, educational level, language skills, cultural peculiarities, and subjective preferences) should be considered at an early stage of the design process and considered in optimizing EMIs further. Coproduction with young service users is essential during these developmental processes [106]. Second, sustained engagement in using digital tools remains a significant challenge [107], which may be addressed through the use of gamification elements, especially in youth [108,109]. However, in this study, the burden associated with app usage was low, and problems with engagement have mainly been reported for stand-alone mHealth apps without components of blended care [110]. Third, in working toward more personalized mHealth apps, more sophisticated methods may be used to inform the timing and context of when intervention components are offered (eg, by using mobile sensing data). A broader range of intervention components delivered for a longer intervention period may help enhance the effects of EMIconpass further and achieve sustainable change in individuals’ daily lives. Fourth, the type of intervention components may be personalized further by assessing the effects of specific intervention components on individuals’ mental health at the individual level. Fifth, it should be further examined whether and, if so, how the therapeutic alliance can be strengthened in light of a limited number of face-to-face sessions [111]. Finally, the number of signals per day triggered by the smartphone was perceived as burdensome by some participants. Thus, future versions of the EMIconpass app may lower the number of signals per day or shorten the number of items per signal [112].

**Conclusions**

Evidence on feasibility and safety and preliminary evidence on the therapeutic effects of the EMIconpass intervention suggest
that translating CFI components into individuals’ daily life through an EMI delivered by an mHealth app may be a promising novel, accessible, and transdiagnostic treatment approach in help-seeking youth by strengthening emotional resilience and directly targeting candidate psychological mechanisms. As an important next step, an exploratory RCT is warranted to demonstrate the feasibility and preliminary evidence of the efficacy of the EMIcompass intervention.

Acknowledgments
This work was supported by the Netherlands Organisation for Scientific Research (grant 451-13-022) and the German Research Foundation (grant 389624707). These funding sources had no further role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. The authors would like to thank Inge Heunen, Danny Deckers, Nele Soons, Christiane Schittek, Shandery Rosalina, and Truda Driesen who helped with recruitment and data collection.

Authors’ Contributions
CR was involved in developing the methodology, formal analysis, data curation, visualization, and writing—original draft, review, and editing. BB was involved in conceptualization, resources, and writing—review and editing. IP was involved in writing—review and editing. KS was involved in obtaining resources, investigation, writing—review and editing—and funding acquisition. AS involved in writing—review and editing. TVA was involved in obtaining resources, investigation, writing—review and editing—and funding acquisition. UR was involved in conceptualization, methodology, formal analysis, writing of the original draft, investigation, resources, further writing—review and editing—supervision, project administration, and funding acquisition.

Conflicts of Interest
None declared.


Abbreviations

- BSI: Brief Symptom Inventory
- CBT: cognitive behavioral therapy
- CF: compassion-focused intervention
- CFI: compassion-focused intervention
- EMA: ecological momentary assessment
- EMI: ecological momentary intervention
- ESM: experience sampling method
- mHealth: mobile health
- PQ: prodromal questionnaire
- RCT: randomized controlled trial

Edited by J Torous; submitted 10.11.20; peer-reviewed by A Beck, E Kleiman; comments to author 21.12.20; revised version received 27.04.21; accepted 25.05.21; published 05.08.21.

Please cite as:
A Compassion-Focused Ecological Momentary Intervention for Enhancing Resilience in Help-Seeking Youth: Uncontrolled Pilot Study
JMIR Ment Health 2021;8(8):e25650
URL: https://mental.jmir.org/2021/8/e25650
doi:10.2196/25650
PMID:
Mental Health and the Perceived Usability of Digital Mental Health Tools Among Essential Workers and People Unemployed Due to COVID-19: Cross-sectional Survey Study

Felicia Mata-Greve¹, PhD; Morgan Johnson¹, MS; Michael D Pullmann¹,², PhD; Emily C Friedman¹,², CPE, MID; Isabell Griffith Fillipo¹,³, BA; Katherine A Comtois¹,⁴, MPH, PhD; Patricia Arean¹,²,³, PhD

¹Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States
²Advanced Laboratories for Accelerating the Reach and Impact of Treatments for Youth and Adults with Mental Illness (ALACRITY) Center, University of Washington, Seattle, WA, United States
³Creating Technological Innovations (CREATIV) Lab, University of Washington, Seattle, WA, United States
⁴Center for Suicide Prevention and Recovery, University of Washington, Seattle, WA, United States

Corresponding Author:
Patricia Arean, PhD
Department of Psychiatry and Behavioral Sciences
University of Washington
1959 NE Pacific Street
Seattle, WA, 98195
United States
Phone: 1 206 616 2129
Email: parean@uw.edu

Abstract

Background: COVID-19 has created serious mental health consequences for essential workers or people who have become unemployed as a result of the pandemic. Digital mental health tools have the potential to address this problem in a timely and efficient manner.

Objective: The purpose of this study was to document the extent of digital mental health tool (DMHT) use by essential workers and those unemployed due to COVID-19, including asking participants to rate the usability and user burden of the DMHT they used most to cope. We also explored which aspects and features of DMHTs were seen as necessary for managing stress during a pandemic by having participants design their own ideal DMHT.

Methods: A total of 2000 people were recruited from an online research community (Prolific) to complete a one-time survey about mental health symptoms, DMHT use, and preferred digital mental health features.

Results: The final sample included 1987 US residents that identified as either an essential worker or someone who was unemployed due to COVID-19. Almost three-quarters of the sample (1479/1987, 74.8%) reported clinically significant emotional distress. Only 14.2% (277/1957) of the sample used a DMHT to cope with stress associated with COVID-19. Of those who used DMHTs to cope with COVID-19, meditation apps were the most common (119/261, 45.6%). Usability was broadly in the acceptable range, although participants unemployed due to COVID-19 were less likely to report user burden with DMHTs than essential workers (t₁₉₈₈.₁ = –3.89, P < .001). Individuals with emotional distress reported higher financial burden for their DMHT than nondistressed individuals (t₆₉.₀ = –3.21, P = .01). When the sample was provided the option to build their own DMHT, the most desired features were a combination of mindfulness/meditation (1271/1987, 64.0%), information or education (1254/1987, 63.1%), distraction tools (1170/1987, 58.9%), symptom tracking for mood and sleep (1160/1987, 58.4%), link to mental health resources (1140/1987, 57.4%), and positive psychology (1131/1986, 56.9%). Subgroups by employment, distress, and previous DMHT use status had varied preferences. Of those who did not use a DMHT to cope with COVID-19, most indicated that they did not consider looking for such a tool to help with coping (1179/1710, 68.9%).

Conclusions: Despite the potential need for DMHTs, this study found that the use of such tools remains similar to prepandemic levels. This study also found that regardless of the level of distress or even past experience using an app to cope with COVID-19, it is possible to develop a COVID-19 coping app that would appeal to a majority of essential workers and unemployed persons.
Health Care System created a free mobile app to help veterans during COVID-19 [8,12]. In anticipation of the need for low-health symptoms and address the mental health care shortage available as potential solutions to decrease stress and mental to allow for better access to health care [11]. DMHTs are also United States, Medicare restrictions on telemedicine were lifted perceived concerns specific to COVID-19. For example, in the such as online resources or mobile phone apps to address perceived concerns specific to COVID-19. For example, in the United States, Medicare restrictions on telemedicine were lifted to allow for better access to health care [11]. DMHTs are also available as potential solutions to decrease stress and mental health symptoms and address the mental health care shortage during COVID-19 [8,12]. In anticipation of the need for low- or no-cost care, organizations such as the Veterans Affairs Health Care System created a free mobile app to help veterans cope with COVID-19. A report from March 2020, as physical distancing began in the United States, found that there was an increased volume of people using these tools [13]. In addition, many organizations and tech companies are turning to DMHTs to support the emotional well-being of frontline health care workers [14].

These recent events lend an important opportunity to learn about the utility of digital mental health to support populations impacted by prolonged pandemic conditions. No research has evaluated the use of DMHTs by two of the most affected populations outside of frontline health care workers and older adults or adults with disability: essential workers and those unemployed due to COVID-19. As identified in several studies, the use of DMHTs tends to be poor, with most people downloading then discontinuing use of these tools in quick succession [15,16]. As Mohr and colleagues [17] have noted, digital mental health service use could be improved if intervention developers better understood what features people felt were important to have, the usability of these tools, and what role these services should have in the context of mental wellness [18-20].

This Study
Considering the need to better understand the mental health challenges faced by essential workers and those unemployed due to COVID-19, the potential long-term effects of the societal challenges imposed by the pandemic, the potential for future pandemics, and the limited information we have on the usability and user burden of DMHTs to cope with the stress of COVID-19, we conducted a study with the following aims:

- **Aim 1:** Document psychological distress through clinically validated measures by the total sample, employment status (ie, unemployed due to COVID-19 and essential workers), and DMHT use (ie, reported using DMHTs to cope with COVID-19, reported not using DMHTs to cope with COVID-19);
- **Aim 2:** Explore DMHT use in response to COVID-19–related stress and differences by employment status and psychological distress (ie, distressed, not distressed);
- **Aim 3:** Assess usability and user burden ratings of DMHTs by total sample, employment status, and psychological distress;
- **Aim 4:** Understand the needs of these at-risk populations by identifying what DMHT features were ranked as most important by employment status, psychological distress, and DMHT use during this time.

Methods
Recruitment
A total of 2000 adults (≥18 years old) were recruited from Prolific Research Platform [21]. Using online research platforms

https://mental.jmir.org/2021/8/e28360
is becoming increasingly popular in behavioral health research due to its affordability, efficiency, access, and reliability [22]. Recent studies highlight that participants recruited from Prolific are more diverse and honest as well as provide higher data quality compared to other popular platforms, such as Amazon Mechanical Turk [22,23]. This national, cross-sectional study collected responses from October 26, 2020, to December 14, 2020. Participants were screened and invited to consent for participation in the anonymous, confidential survey online. Each participant was paid $3. The research was approved by the University of Washington’s institutional review board.

Measures
Measures were selected and created to maximize participant engagement and reduce respondent burden. The investigative team reviewed brief measures of constructs of interest and gave preference to longer measures where no reliable or valid brief measure was available.

Inclusion Screening
Participants must have been ≥18 years old, speak English, and self-reported as either an essential worker during COVID-19 or unemployed or furloughed due to COVID-19. They also had the opportunity to indicate their current job (if an essential worker) or past job (if an unemployed worker). Participants were excluded if they were under 18 years of age, did not speak English, had no access to a mobile device (e.g., smartphone or tablet), did not report being an essential worker or unemployed due to COVID-19, or lived outside of the United States.

Bad-Actor Screening
Even with the best safeguards in place, online recruitment can sometimes result in the accidental inclusion of individuals who participate in bad faith to accumulate monetary incentives (“bad actors”) [24]. We instituted the procedures explained below to identify potential bad actors.

The first was to use research platforms (described above) that conduct their own extensive participant vetting. These procedures include but are not limited to: (1) every account needing a unique non-VOIP (voice over IP) phone number to verify, (2) restricting signups based on IP address and internet service provider, (3) limiting the number of accounts that can use the same IP address and machine to prevent duplicate accounts, (4) limiting the number of unique IP addresses per study, and (5) unique payment accounts (e.g., PayPal) for each participant account. For example, in order to have 2 participant accounts that receive payment from Prolific, a participant would need to have 2 PayPal accounts. Payment accounts, such as PayPal, have steps to prevent duplicate accounts, such as analyzing internal data to monitor for patterns of unusual use [25].

The second method involved the use of an attention check built into our survey [26]. This method consisted of one question where participants were given this instruction: “To confirm you are paying attention, please select ‘strongly disagree’” and then choices between strongly agree to strongly disagree were provided.

The third method involved the review of open-ended responses to screen out bot-like communication, repetitive, and nonsensical responses. Each of these methods confirmed that the final sample in this study could be qualified as comprising “good actors.”

Demographics
Participants completed a questionnaire about demographics, which collected information about age, race, ethnicity, gender identity, sexual orientation, marital status, education, employment status, income, and living situation.

Mental Health and Possible Substance Use Disorder
Participants completed the 2-item Patient Health Questionnaire (PHQ-2) [27], the 2-item Generalized Anxiety Disorder (GAD-2) [28], and the Cut-Annoyed-Guilty-Eye Adapted to Include Drugs (CAGE-AID) [29]. The PHQ-2 and GAD-2 have good sensitivity and specificity with sensitivity to change over time in comparison to the PHQ-9 and GAD-7 [28-30]. The CAGE-AID demonstrates good sensitivity and poor specificity for substance use disorders. As a result, individuals who scored beyond the cut-off on the CAGE-AID (≥1) were categorized as a possible case of substance use disorder, in accordance with the National HIV Curriculum [29,31].

Suicidal Behaviors
Suicidal behaviors were measured using the Suicide Behaviors Questionnaire–Revised (SBQ-R) [32], a 4-item self-report measure that assesses suicide attempts, ideation, communication, and intent in one’s lifetime. If the total score is greater than or equal to 7, the score is deemed to have good sensitivity and specificity for identifying individuals at risk for suicidal behaviors in a nonpsychiatric general adult population. Given some limitations of the SBQ-R, a single validated item (ie, “Have you attempted to kill yourself?”) was added. The addition of this item provides further accuracy and classification of individuals at risk of suicide [33].

Psychological Distress
Participants were placed in the “distressed” category if they endorsed one or more of the clinical cut-offs, which included ≥3 on the PHQ-2 [27], ≥3 on the GAD-2 [28], ≥1 on the CAGE-AID, ≥7 [29,31] on SBQ-R [32], or reported a history of a suicide attempt [33].

DMHT Questionnaire
This questionnaire was developed by the research team with expertise in digital mental health (author PA). The measure was tested for face validity, understandability, and respondent burden among the internal group. The questionnaire consisted of three distinct tasks: use of DMHTs during COVID-19, usability and burden of DMHTs during COVID-19, and design of an ideal DMHT for COVID-19, which are described below.

Use of DMHTs
All participants were asked whether they have used an app to cope with stress associated with COVID-19. If the participant responded yes, they were asked to list which apps they used, and if they used more than one, to list the app they used the most to cope with COVID-19. Participants were then asked to...
rate the app that they used most frequently in terms of features they liked, features they did not like, and then on the app’s usability and user burden. If participants did not report using an app to cope with COVID-19 stress, they were asked to provide reasons for why they did not use an app (Figure 1).

**Usability**

Usability was measured with the System Usability Scale (SUS) [34], a 10-item measure that examines the usability of a particular intervention. The scale assesses a system’s likability, learnability, complexity, need for technical support, system integration, and efficiency. The SUS is the industry standard for measuring the usability of a variety of digital tools and systems and has normative data to allow for cross system and app comparisons, even between those that are outwardly very dissimilar to one another [35].

**User Burden**

User burden was measured using the 20-item Use Burden Scale (UBS) [36]. This scale creates five subscales to assess different types of user burden: difficulty of use (“this app demands too much mental effort”), physical demands (“use of this app is too physically demanding”), time and social burden (“I spend too much time using this app”; “using this app has a negative impact on my social life”), mental and emotional burden (“this app presents too much information at once”), and privacy and financial burden (“the value of the app is not worth the cost for me”). This measure was developed in order to assess the adoption, retention, and experience of various technologies with the ability to compare and calibrate burden across different tools. User burden is linked to app retention and has been used in the context of mobile app research [37].

**Design of a COVID-19 App**

All participants, regardless of whether they reported app use for stress associated with COVID-19, were asked which features they thought would be helpful to include in an app for coping with COVID-19 (ie, information or education, meditation/mindfulness, symptom tracking, brain games, distraction tools, gratitude exercises, links to resources, chatbot, or tips to cope with COVID-19) on a scale from 0="not at all important” to 9=”very important.” This method of asking opinions of those who do and do not use digital technology, particularly when the needs of a given population are unknown, is commonly used in app development. The opinions of people familiar and unfamiliar with apps are needed to design a digital tool with the broadest reach [38].

After indicating which features participants preferred in an app to cope with COVID-19, they were then asked to build their own app, by selecting from a preset list of features and then adding their own desired features that were not previously listed. The app feature list was created using premade categories from One Mind Psybgeruide [39], a nonprofit tool that reviews digital mental health tools for consumers, and M-Health Index and Navigation Database (MIND) [40] (see Multimedia Appendix 1 for the full survey).

**Statistical Analysis**

To describe the sample, we ran crosstabulations (with chi-square tests or Fisher exact tests) and independent samples t tests to examine possible differences in the demographic and descriptive variables by employment status (ie, unemployed vs essential worker groups) and DMHT use (ie, DMHT user vs non–DMHT user). For variables with multiple discrete categories (eg, education), if these analyses indicated a significant omnibus chi-square test, we examined standardized residuals to identify which categories were responsible for the omnibus significant difference, and reported on all categories with absolute value standardized residuals greater than 2.

For the first aim, descriptive statistics were used to document the frequencies and means of the psychological distress composite among the entire sample and stratified by employment status. We also compared those who reported using an app to cope with COVID-19 to those who reported not using an app to cope with COVID-19. Specific reports on depression, anxiety, possible substance use disorder, suicidal behavior, and history of suicide attempt may be found in Multimedia Appendix 2.

For the second aim, we calculated frequencies and differences in DMHT use for the whole sample, between essential workers and those unemployed and between those reporting distress and no distress.

For the third aim, we computed means and SDs to examine DMHT ratings from the SUS and the UBS only for those who reported using a DMHT to cope with COVID-19. Differences across the top 3 apps were assessed using an ANOVA (analysis of variance). For the sample that did not report using a DMHT to cope during COVID-19, we provided the reasons for not using a DMHT and the frequency by which those reasons were endorsed in the sample.

For the fourth aim, we computed frequencies and central tendencies of the data to assess preferred DMHT components for the whole sample and compared these findings first between essential workers and those unemployed, then between distressed and nondistressed subsamples, and finally between those who reported having used a DMHT and those who did not.

The aims described above that examined significant differences by employment, distress, or DMHT use status were assessed using chi-square tests, Fisher exact tests, or independent samples t tests. All statistical analyses were performed with SAS version 9.4 (SAS Institute Inc.). To adjust for increased type I error rates due to multiple tests, we applied the Benjamini-Hochberg procedure, which applies the acceptable fraction of tests that may be erroneously statistically significant, deemed the “false discovery rate” [41,42]. We applied a false discovery rate (Q) of 10% to 119 statistical tests.

Open-ended responses from the DMHT survey for categories (ie, “What app did you try? If you tried more than one app, please pick the one you liked the most”) and app features listed during the create-your-own-app survey were qualitatively coded. Like Rubanovich et al [43], the first author (FM-G) referenced the Apple App Store and Google Play to verify spelling and
DMHT titles. As an example, *Calm, CALM, Calm App, Calm,* and *Camh* were all coded as “Calm.” If a DMHT was unable to be identified via Google Play, Apple App Store, or an internet search, or the participant response was undecipherable (eg, “IDK,” “NA”), it was categorized as missing (n=18).

Categorization of DMHTs was completed by authors FM-G and MJ. Informed by a modified grounded theory approach [44], each response was reviewed in order to identify meaningful units of information. Responses were compared with one another and grouped based on common responses until categories were identified. If the authors were unfamiliar with a DMHT, they read descriptions and reviews of the DMHT to determine its main feature. Some participants described DMHTs instead of names. In these cases, the response was coded for a DMHT category, but not for a specific DMHT title. As an example, the following responses, “I used a few meditation apps and one about CBT,” “mindfulness app,” and “meditation app” were coded into the mindfulness/meditation category. Categories and definitions were informed by Psyberguide, MIND, and experience working with digital mental health researchers. An identical process was conducted to code desired app features.

Data Exclusion and Cleaning
Duplicate cases were identified and removed. Missingness accounted for less than 5% of the data evaluated item by item. Measures were scored unless all items were missing. As an exception, PHQ-2, GAD-2, and CAGE-AID required all items to be answered to attain a final score.

Results
Sample Description
A total of 2485 participants completed the initial screener. Of this, 598 (23.7%) observations were deleted due to missing IDs, duplicate responses, “bad actors,” or not meeting inclusion criteria. The final analytic sample (Table 1) consisted of 1987 adults with 1013 (50.9%) participants reporting unemployment due to COVID-19 and 974 (49.0%) identifying as an essential worker during COVID-19. The most common open-ended responses for jobs among essential workers included education, customer service or retail, management, information technology (IT), health care, pharmacy, delivery or postal work, and food service (eg, cashiers, servers, restaurant workers, grocery store workers). Although we sampled throughout the United States, compared to the US census, the majority of the overall sample was European American (1538/1987, 77.4%, compared to the US census figure of 60%), with a somewhat higher representation of Asian Americans (238/1987, 12.0% vs 5% US census) and a lower representation of African Americans (172/1987, 8.7% vs 13% US census) and Latinx Americans (212/1956, 10.8% vs 18% US census) [45]. The sample was almost split evenly between male and female (female: 1027/1987, 52.2%).

Compared to the essential workers, the unemployed group had significantly more people who identified as being: Hispanic or Latinx, or an unlisted race; younger; any gender other than male; any sexuality other than straight; and never married. The group comprised significantly less White individuals. Of note, there were almost twice as many in the “single or never married” category than what would be expected compared to the US census data [46]; however, our sample was relatively young (ie, early 30s) compared to the US population [47]. Additionally, there were socioeconomic differences across groups. Compared to the essential workers, the unemployed group had significantly more individuals with lower education, less income, and lived somewhere other than a house or apartment.

Compared to participants that did not use a DMHT to cope with COVID-19 stress, DMHT users had a significantly higher proportion of individuals who identified as transgender and a lower proportion of individuals who identified as women or men. DMHT users were more likely to be married compared to non–DMHT users. In terms of socioeconomic differences, DMHT users had a significantly smaller proportion of individuals with lower levels of education and a higher percentage of individuals with higher education compared to non–DMHT users. Finally, compared to non–DMHT users, DMHT users were less likely to live in a house and more likely to live in an apartment.
Table 1. Sample characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Unemployed (n=1013)</th>
<th>Essential worker (n=974)</th>
<th>P value</th>
<th>Non–DMHT user (n=1680)</th>
<th>DMHT user (n=277)</th>
<th>P value</th>
<th>Total (N=1987)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race (not mutually exclusive), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian American</td>
<td>129 (12.7)</td>
<td>109 (11.2)</td>
<td>.29a</td>
<td>207 (12.3)</td>
<td>29 (10.5)</td>
<td>.38a</td>
<td>238 (12.0)</td>
</tr>
<tr>
<td>European American/White</td>
<td>763 (75.3)</td>
<td>775 (79.6)</td>
<td>.02a</td>
<td>1300 (77.4)</td>
<td>225 (81.2)</td>
<td>.15a</td>
<td>1538 (77.4)</td>
</tr>
<tr>
<td>African American/Black</td>
<td>95 (9.4)</td>
<td>77 (7.9)</td>
<td>.24a</td>
<td>153 (9.1)</td>
<td>16 (5.8)</td>
<td>.07a</td>
<td>172 (8.7)</td>
</tr>
<tr>
<td>Hawaiian/Pacific Islander</td>
<td>8 (0.8)</td>
<td>3 (0.3)</td>
<td>.23b</td>
<td>11 (0.7)</td>
<td>0 (0)</td>
<td>.38b</td>
<td>11 (0.6)</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>24 (2.4)</td>
<td>24 (2.5)</td>
<td>.89b</td>
<td>36 (2.1)</td>
<td>9 (3.2)</td>
<td>.26b</td>
<td>48 (2.4)</td>
</tr>
<tr>
<td>Unlisted</td>
<td>52 (5.1)</td>
<td>22 (2.3)</td>
<td>&lt;.001a</td>
<td>63 (3.8)</td>
<td>11 (4.0)</td>
<td>.86a</td>
<td>74 (3.7)</td>
</tr>
<tr>
<td>Ethnicity, (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic/Latinx</td>
<td>128 (12.9)</td>
<td>84 (8.7)</td>
<td>.38a</td>
<td>179 (10.8)</td>
<td>30 (11.0)</td>
<td>.02a</td>
<td>212 (10.8)</td>
</tr>
<tr>
<td>Not Hispanic/Latinx</td>
<td>863 (87.1)</td>
<td>881 (91.3)</td>
<td></td>
<td>1483 (89.2)</td>
<td>243 (89.0)</td>
<td>.01a</td>
<td>1744 (89.2)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td>&lt;.001c</td>
<td></td>
<td></td>
<td>.92c</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>30.4 (11.1)</td>
<td>33.3 (9.9)</td>
<td></td>
<td>31.8 (10.8)</td>
<td>31.9 (9.7)</td>
<td></td>
<td>31.9 (10.6)</td>
</tr>
<tr>
<td>Range</td>
<td>18.0-73.0</td>
<td>18.0-78.0</td>
<td></td>
<td>18.0-78.0</td>
<td>18.0-73.0</td>
<td></td>
<td>18.0-78.0</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001b</td>
<td></td>
<td>&lt;.001b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>573 (57.4)</td>
<td>454 (46.9)</td>
<td></td>
<td>853 (51.1)</td>
<td>166 (60.1)</td>
<td></td>
<td>1027 (52.2)</td>
</tr>
<tr>
<td>Men</td>
<td>384 (38.4)</td>
<td>499 (51.5)</td>
<td></td>
<td>774 (46.3)</td>
<td>96 (34.8)</td>
<td></td>
<td>883 (44.9)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>35 (3.5)</td>
<td>12 (1.2)</td>
<td></td>
<td>38 (2.3)</td>
<td>9 (3.3)</td>
<td></td>
<td>47 (2.4)</td>
</tr>
<tr>
<td>Transgender</td>
<td>3 (0.3)</td>
<td>2 (0.2)</td>
<td></td>
<td>2 (0.1)</td>
<td>3 (1.1)</td>
<td></td>
<td>5 (0.3)</td>
</tr>
<tr>
<td>Unlisted</td>
<td>4 (0.4)</td>
<td>1 (0.1)</td>
<td></td>
<td>3 (0.2)</td>
<td>2 (0.7)</td>
<td></td>
<td>5 (0.3)</td>
</tr>
<tr>
<td>Sexuality, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
<td></td>
<td>.30a</td>
<td></td>
</tr>
<tr>
<td>Heterosexual/straight</td>
<td>681 (69.0)</td>
<td>802 (82.9)</td>
<td></td>
<td>1276 (76.6)</td>
<td>192 (71.6)</td>
<td></td>
<td>1483 (75.9)</td>
</tr>
<tr>
<td>Gay/lesbian/homosexual</td>
<td>69 (7.0)</td>
<td>41 (4.2)</td>
<td></td>
<td>89 (5.3)</td>
<td>20 (7.5)</td>
<td></td>
<td>110 (5.6)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>189 (19.1)</td>
<td>104 (10.8)</td>
<td></td>
<td>243 (14.6)</td>
<td>45 (16.8)</td>
<td></td>
<td>293 (15.0)</td>
</tr>
<tr>
<td>Unlisted</td>
<td>48 (4.9)</td>
<td>20 (2.1)</td>
<td></td>
<td>57 (3.4)</td>
<td>11 (4.1)</td>
<td></td>
<td>68 (3.5)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
<td></td>
<td>.02a</td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>737 (73.8)</td>
<td>500 (51.9)</td>
<td></td>
<td>1065 (63.8)</td>
<td>156 (57.6)</td>
<td></td>
<td>1237 (63.0)</td>
</tr>
<tr>
<td>Widowed</td>
<td>8 (0.8)</td>
<td>5 (0.5)</td>
<td></td>
<td>13 (0.8)</td>
<td>0 (0)</td>
<td></td>
<td>13 (0.7)</td>
</tr>
<tr>
<td>Married</td>
<td>177 (17.7)</td>
<td>402 (41.7)</td>
<td></td>
<td>473 (28.3)</td>
<td>101 (37.3)</td>
<td></td>
<td>579 (29.5)</td>
</tr>
<tr>
<td>Separated</td>
<td>15 (1.5)</td>
<td>7 (0.7)</td>
<td></td>
<td>21 (1.3)</td>
<td>1 (0.4)</td>
<td></td>
<td>22 (1.1)</td>
</tr>
<tr>
<td>Divorced</td>
<td>61 (6.1)</td>
<td>50 (5.2)</td>
<td></td>
<td>98 (5.9)</td>
<td>13 (4.8)</td>
<td></td>
<td>111 (5.7)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
</tr>
<tr>
<td>High school graduate (or equivalent) or less</td>
<td>154 (15.3)</td>
<td>77 (7.9)</td>
<td></td>
<td>215 (12.8)</td>
<td>11 (4.0)</td>
<td></td>
<td>231 (11.7)</td>
</tr>
<tr>
<td>Some college</td>
<td>367 (36.5)</td>
<td>192 (19.7)</td>
<td></td>
<td>480 (28.6)</td>
<td>74 (26.7)</td>
<td></td>
<td>559 (28.3)</td>
</tr>
<tr>
<td>Trade/technical/vocational training/associate degree</td>
<td>125 (12.4)</td>
<td>108 (11.1)</td>
<td></td>
<td>209 (12.4)</td>
<td>22 (7.9)</td>
<td></td>
<td>233 (11.8)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>283 (28.2)</td>
<td>353 (36.3)</td>
<td></td>
<td>540 (32.1)</td>
<td>92 (33.2)</td>
<td></td>
<td>636 (32.2)</td>
</tr>
<tr>
<td>Higher education (master’s, professional, or doctorate degree)</td>
<td>76 (7.6)</td>
<td>243 (25.0)</td>
<td></td>
<td>236 (14.0)</td>
<td>78 (28.2)</td>
<td></td>
<td>319 (16.1)</td>
</tr>
</tbody>
</table>
Table 2 reports psychological distress (see the Measures section for calculation of the composite score) for the whole sample with stratification by employment status and DMHT-use status. We found that almost three-quarters of the sample fell into the “distressed” category (1479/1976, 74.8%), meaning they had scores at or above the clinical cut-off for at least one of the following: depression (PHQ-2), anxiety (GAD-2), risk for substance use disorder (CAGE-AID), risk for suicidal behaviors (SBQ-R), and history of suicide attempt. The unemployed group was more likely to be distressed than the essential worker group (815/1013, 81.2% vs 664/974, 68.3%; \( \chi^2 = 43.40, P < .001; \) Table 2). DMHT users were significantly more likely to be distressed compared to non–DMHT users (236/277, 85.2% vs 1234/1680, 73.5%; \( \chi^2 = 17.55, P < .001; \) Table 2). Table S1 in Multimedia Appendix 3 provides a further breakdown of depression, anxiety, risk for substance use disorder, risk for suicidal behaviors, and history of suicide attempt by total sample, employment status, and DMHT-use status.

Table 2. Psychological distress stratified by employment status and digital mental health tool (DMHT) use.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unemployed (n=1013)</th>
<th>Essential worker (n=974)</th>
<th>( P ) value</th>
<th>Non–DMHT user (n=1680)</th>
<th>DMHT user (n=277)</th>
<th>( P ) value</th>
<th>Total (N=1987)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychological distress, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nondistressed</td>
<td>189 (18.8)</td>
<td>308 (31.7)</td>
<td>&lt;.001(^{a,b})</td>
<td>446 (26.5)</td>
<td>41 (14.8)</td>
<td>&lt;.001(^{a,b})</td>
<td>497 (25.2)</td>
</tr>
<tr>
<td>Distressed</td>
<td>815 (81.2)</td>
<td>664 (68.3)</td>
<td></td>
<td>1234 (73.5)</td>
<td>236 (85.2)</td>
<td></td>
<td>1479 (74.8)</td>
</tr>
</tbody>
</table>

\(^{a}\)Chi-square test.
\(^{b}\)Fisher exact test.

Aim 2: Explore DMHT Use in Response to COVID-19

Of the 1957 participants who responded, 277 (14.2%) reported using a DMHT to cope with stress associated with COVID-19. There was no significant difference in the proportion of participants who used a DMHT in the unemployed (137/1013, 13.5%) and essential worker (140/974, 14.4%) groups (\( \chi^2 = 0.25, P = .62 \)). Distressed individuals (236/1470, 16.1%) were significantly more likely to use a DMHT app compared to nondistressed individuals (41/487, 8.4%; \( \chi^2 = 17.55, P < .001 \)).
Table 3. Categories of digital mental health tools (DMHTs).

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Participants, n (%)^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meditation/mindfulness</td>
<td>A DMHT offering primarily meditation or mindfulness (eg, Calm, Headspace)</td>
<td>119 (45.6)</td>
</tr>
<tr>
<td>Virtual therapy or contact with a virtual provider</td>
<td>A DMHT offering primarily virtual therapy via text, phone, or video, or appointments with a physician (eg, BetterHelp, Sanvello)</td>
<td>25 (9.6)</td>
</tr>
<tr>
<td>Chat feature</td>
<td>The main feature was a chat function for one-on-one chats with a peer or chatbot, group chats, or connecting with others in an organized forum (eg, Woebot, Wysa)</td>
<td>21 (8.1)</td>
</tr>
<tr>
<td>Health</td>
<td>Tools that offer education or tips to promote healthy habits with exercise, nutrition, physical health, or sleep (eg, DownDog)</td>
<td>20 (7.7)</td>
</tr>
<tr>
<td>COVID-19 contact tracing</td>
<td>A DMHT with information related to local COVID-19 cases, rates of infection, and information about symptoms or testing (eg, Contact Tracing)</td>
<td>13 (5.0)</td>
</tr>
<tr>
<td>Entertainment and distraction</td>
<td>A DMHT with entertainment, which may include movies, music, games, GIFs, memes, or other forms of entertainment (eg, Among Us, Music app)</td>
<td>12 (4.6)</td>
</tr>
<tr>
<td>Social media</td>
<td>A social media platform (eg, TikTok, Reddit)</td>
<td>10 (3.8)</td>
</tr>
<tr>
<td>Symptom tracking</td>
<td>A DMHT that allows users to monitor symptoms or daily activities (eg, eMoods, The Pattern)</td>
<td>10 (3.8)</td>
</tr>
<tr>
<td>COVID-19 coping</td>
<td>A DMHT providing emotional coping skills and education in the context of COVID-19 stressors (eg, COVID Coach)</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>Positive psychology</td>
<td>A DMHT with gratitude exercises or methods to promote positivity, such as daily verses, positive thoughts, uplifting stories, or uplifting quotes (eg, InnerHour)</td>
<td>7 (2.7)</td>
</tr>
<tr>
<td>Finance</td>
<td>A DMHT with resources for financial decisions, financial decision-making, or spending tips (eg, Yes, Pacific)</td>
<td>7 (2.7)</td>
</tr>
<tr>
<td>Journal</td>
<td>A DMHT with primarily writing or journaling features (eg, Day One, Iona)</td>
<td>4 (1.5)</td>
</tr>
<tr>
<td>News</td>
<td>Information about international or national occurrences (eg, WHO Info)</td>
<td>3 (1.2)</td>
</tr>
<tr>
<td>Crisis</td>
<td>Using a DMHT to manage crisis or safety (eg, suicide)</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Language learning</td>
<td>Using a DMHT in order to practice or learn a new language</td>
<td>1 (0.4)</td>
</tr>
</tbody>
</table>

^a A total of 18 responses were coded as “missing” due to being indecipherable or unidentifiable; percentages do not reflect missingness.

**Employment Status**

The leading entries by the unemployed sample were 3 meditation apps: Calm (26/131, 19.8%), Headspace (22/131, 16.8%), and Insight Timer (7/131, 5.3%). The most common DMHT categories among individuals unemployed due to COVID-19 were meditation (70/131, 53.4%), virtual therapy or DMHTs that facilitated virtual contact with a mental health provider (11/131, 8.4%), and DMHTs with a chatbot (11/131, 8.4%). The most frequently reported DMHTs by the essential worker sample were Headspace (16/130, 12.3%), Calm (15/130, 11.5%), and COVID Coach (8/130, 6.2%). By category, essential workers reported using mostly meditation (49/130, 37.7%), DMHTs with virtual therapy or contact with a virtual provider (14/130, 10.8%), health DMHTs (12/130, 9.4%), and COVID-19 contact tracing (12/130, 9.4%).

**Distress Status**

Similarly, the leading entries by the distressed sample were 2 meditation apps, Calm (33/223, 14.8%) and Headspace (32/223, 14.3%), followed by BetterHelp (10/223, 4.5%). Most of the distressed sample used meditation (100/223, 44.8%), virtual therapy or contact with a virtual provider (24/223, 10.8%), and DMHTs with a chat feature (19/223, 8.5%). The most frequently reported DMHTs by the nondistressed group were Calm, (8/38, 21.1%), Headspace (6/38, 15.8%), and COVID Coach (2/38 5.3%). Among the individuals in the nondistressed group, the most frequently used app categories were meditation (9/38, 50%), COVID-19 contact tracing (4/38, 10.5%), and social media (3/38, 7.9%).

Further comparisons of app categories by employment and distress statuses may be found in Table S2 (Multimedia Appendix 3).

**Reasons for Lack of Use**

Most of the sample (1710/1957, 85.9%) reported that they did not use a DMHT to cope with COVID-19. The primary reasons for not using a DMHT to cope with COVID-19 were (1) not thinking to look for an app (1179/1710, 68.9%), (2) not thinking apps would help them (605/1710, 35.4%), and (3) having other ways of coping (421/1710, 24.6%). Table S3 in Multimedia Appendix 3 lists all reasons for lack of use. These top 3 responses were endorsed by all subgroups.

There were differences that emerged by employment status and distress status. Compared to essential workers, those who were unemployed were more likely to use DMHTs for meditation, virtual therapy, and social media. Distressed individuals were more likely to use DMHTs for meditation, virtual therapy, and COVID-19 contact tracing.

---

https://mental.jmir.org/2021/8/e28360
unemployed due to COVID-19 were more likely to report not thinking to look for a DMHT (629/876, 71.8% vs 550/834, 65.9%; \( \chi^2 = 6.84, P = .009 \)) and not having money to spend on a data plan to use a DMHT (112/876, 12.8% vs 54/834, 6.5%; \( \chi^2 = 19.41, P < .001 \)).

Compared to the nondistressed group, distressed individuals were more likely to not think to look for an app (293/456, 64.3% vs 886/1243, 71.3%; \( \chi^2 = 7.75, P = .005 \)), to not think apps would help them (142/456, 31.1% vs 463/1243, 37.2%; \( \chi^2 = 5.43, P = .02 \)), to prefer working with a professional (32/456, 7.0% vs 191/1243, 15.4%; \( \chi^2 = 20.39, P < .001 \)), to not have money to spend on a data plan to use apps (25/456, 5.5% vs 141/1243, 11.3%; \( \chi^2 = 13.00, P < .001 \)), and to not find an app that was relevant to their needs (19/456, 4.2% vs 103/1243, 8.3%; \( \chi^2 = 8.50, P = .004 \)). However, compared to nondistressed individuals, distressed workers were less likely to state that having another way of coping was the reason for why they did not use a DMHT (281/1243, 22.6% vs 140/456, 30.7%; \( \chi^2 = 11.73, P < .001 \)).

**Aim 3: Assess DMHT Usability and User Burden**

Data for the following analyses were taken from the 277 participants who reported using a DMHT to cope with COVID-19. Individuals who did not report using a DMHT to cope with COVID-19 did not complete the SUS or UBS (Figure 1).

### Employment Status

As shown in Table 4, compared to the essential workers, those who were unemployed due to COVID-19 reported significantly less user burden when using DMHTs (mean 13.69, SD 17.76 vs mean 7.23, SD 8.24; \( t_{198.1} = -3.89, P < .001 \)). Specifically, those who were unemployed rated their selected DMHT as being significantly less difficult to use (mean 2.77, SD 4.02 vs mean 1.53, SD 2.15; \( t_{214.5} = 3.20, P < .002 \)), and having less physical burden (mean 1.54, SD 2.89 vs mean 0.43, SD 1.37; \( t_{198.3} = 4.06, P < .001 \)), time and social burden (mean 2.60, SD 4.00 vs mean 1.07, SD 2.15; \( t_{215.1} = 3.95, P < .001 \)), mental and emotional burden (mean 2.46, SD 3.92 vs mean 1.07, SD 2.08; \( t_{213.4} = 3.69, P < .001 \)), and privacy burden (mean 2.30, SD 3.16 vs mean 1.25, SD 2.14; \( t_{245.1} = 3.25, P < .001 \)). The conditions did not differ for reports of financial burden (mean 2.04, SD 2.33 vs mean 1.88, SD 2.47; \( t_{272} = -0.53, P = .59 \)). In addition, there was no significant difference in ratings of usability between unemployed individuals (mean 76.96, SD 16.21) and essential workers (mean 74.32, SD 17.01; \( t_{271} = -1.31, P = .19 \)).

https://mental.jmir.org/2021/8/e28360
### Table 4. User burden and system usability stratified by workers and psychological distress.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unemployed (n=137)</th>
<th>Essential worker (n=140)</th>
<th>P value</th>
<th>Nondistressed (n=41)</th>
<th>Distressed (n=236)</th>
<th>P value</th>
<th>Total (N=277)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall burden</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>140</td>
<td>.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>41</td>
<td>233</td>
<td>.30&lt;sup&gt;b&lt;/sup&gt;</td>
<td>274</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>7.2 (8.2)</td>
<td>13.7 (17.8)</td>
<td></td>
<td>8.4 (13.8)</td>
<td>10.9 (14.4)</td>
<td>10.5 (14.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Difficulty of use</strong></td>
<td></td>
<td></td>
<td>.002&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>140</td>
<td>.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>41</td>
<td>233</td>
<td>.52&lt;sup&gt;a&lt;/sup&gt;</td>
<td>274</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.5 (2.2)</td>
<td>2.8 (4.0)</td>
<td></td>
<td>1.8 (3.8)</td>
<td>2.2 (3.2)</td>
<td>2.2 (3.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Physical burden</strong></td>
<td></td>
<td></td>
<td>.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.99&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>139</td>
<td></td>
<td>40</td>
<td>233</td>
<td></td>
<td>273</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.4 (1.4)</td>
<td>1.5 (2.9)</td>
<td></td>
<td>0.8 (2.0)</td>
<td>1.0 (2.4)</td>
<td>1.0 (2.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Social and time burden</strong></td>
<td></td>
<td></td>
<td>.001&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.80&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>140</td>
<td></td>
<td>41</td>
<td>233</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.1 (2.2)</td>
<td>2.6 (4.0)</td>
<td></td>
<td>1.9 (3.3)</td>
<td>1.9 (3.3)</td>
<td>1.9 (3.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Mental and emotional burden</strong></td>
<td></td>
<td></td>
<td>.001&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.80&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>140</td>
<td></td>
<td>41</td>
<td>233</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.1 (2.1)</td>
<td>2.5 (3.9)</td>
<td></td>
<td>1.7 (3.3)</td>
<td>1.8 (3.2)</td>
<td>1.8 (3.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Privacy burden</strong></td>
<td></td>
<td></td>
<td>.001&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.22&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>140</td>
<td></td>
<td>41</td>
<td>233</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.2 (2.1)</td>
<td>2.3 (3.2)</td>
<td></td>
<td>1.3 (2.6)</td>
<td>1.9 (2.8)</td>
<td>1.8 (2.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Financial burden</strong></td>
<td></td>
<td></td>
<td>.59&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.002&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>140</td>
<td></td>
<td>41</td>
<td>233</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.9 (2.5)</td>
<td>2.0 (2.3)</td>
<td></td>
<td>1.1 (1.8)</td>
<td>2.1 (2.5)</td>
<td>2.0 (2.4)</td>
<td></td>
</tr>
<tr>
<td><strong>System Usability Score</strong></td>
<td></td>
<td></td>
<td>.19&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.96&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Count, n</td>
<td>134</td>
<td>139</td>
<td></td>
<td>40</td>
<td>233</td>
<td></td>
<td>273</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>77.0 (16.2)</td>
<td>74.3 (17.0)</td>
<td></td>
<td>75.5 (17.6)</td>
<td>75.6 (16.5)</td>
<td>75.6 (16.6)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Equal variance two-sample t test.

<sup>b</sup> Unequal variance two-sample t test.

<sup>c</sup> P values <.05 and less than the Benjamini-Hochberg critical value were considered to be statistically significant.

### Distress Status

As shown in Table 4, there was no difference in reported DMHT burden between the distressed and nondistressed subsamples (mean 10.91, SD 14.37 vs mean 8.41, SD 13.82; t<sub>272</sub>=-1.03, P=.30) or in overall usability (mean 75.63, SD 16.52 vs mean 75.50, SD 17.59; t<sub>271</sub>=-0.05, P=.96). Likewise, we found no difference between groups in types of burden (Table 4). The one exception was that distressed individuals reported higher financial burden for their selected DMHT than nondistressed individuals (mean 2.12, SD 2.46 vs mean 1.07, SD 1.81; t<sub>69.0</sub>=-3.21, P=.01).

Finally, we explored the user burden and usability ratings of the three most used apps (ie, Calm, Headspace, and BetterHelp; shown in Table S4 in Multimedia Appendix 3). There were no statistically significant differences among the apps in terms of the total SUS, total UBS, and UBS subscales, except for privacy burden (Calm: mean 1.54, SD 2.82 vs Headspace: mean 0.50, SD 1.03 vs BetterHelp: mean 2.00, SD 2.14; F<sub>2.87</sub>=3.25, P=.04).

### Aim 4: Identify Important DMHT Features

#### Total Sample

The sample reported the following top-rated features for DMHTs: (1) information or education (mean 6.09, SD 2.66); (2) mindfulness or meditation tools (mean 6.06, SD 2.59); (3) link to resources, counseling, or crisis support (mean 5.93, SD 2.80); and (4) tools to focus on positive events and influences in life (mean 5.88, SD 2.46).

Participants also had the option to write in what DMHT features they felt were important to include but were not provided in the list of options. The top suggested features among the 764 responses were the ability to chat with a mental health professional.
professional, support personnel, or peer (n=57); entertainment and distraction (n=39); and positive psychology (n=29). The feature “entertainment and distraction” was defined as “different forms of entertainment such as music, movies, movie clips, GIFs, memes, games, or other forms of distraction.” Additionally, participants reported wanting regularly occurring (ie, daily) gratitude exercises or activities to promote positivity, such as verses, quotes, and uplifting or hopeful stories, which we categorized as “positive psychology” features. Example responses included: “give positive messages in the morning or something like that,” “daily gratitude,” and “a good news section… I don’t want to be told COVID-19 isn’t a problem. I want to know what hope there is.”

When provided the option to build their own app, the sample most frequently endorsed the following features: mindfulness/meditation (1271/1987, 64.0%), information or education (1254/1987, 63.1%), and distraction tools (1170/1987, 58.9%) (Table 5).

Table 5. Digital mental health tool (DMHT) features stratified by worker status and psychological distress.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Unemployed (n=1013), n (%)</th>
<th>Essential worker (n=974), n (%)</th>
<th>P value</th>
<th>Nondistressed (n=497), n (%)</th>
<th>Distressed (n=1479), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mindfulness/meditation</td>
<td>687 (67.8)</td>
<td>584 (60.0)</td>
<td>&lt;.001&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>305 (61.4)</td>
<td>966 (65.3)</td>
<td>.11&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Information or education</td>
<td>636 (62.8)</td>
<td>618 (63.4)</td>
<td>.76&lt;sup&gt;a&lt;/sup&gt;</td>
<td>327 (65.8)</td>
<td>927 (62.7)</td>
<td>.21&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Distraction tools (drawing, puzzles, music)</td>
<td>630 (62.2)</td>
<td>540 (55.4)</td>
<td>.002&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>276 (55.5)</td>
<td>894 (60.4)</td>
<td>.05&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Symptom tracking (tracking sleep or mood)</td>
<td>605 (59.7)</td>
<td>555 (57.0)</td>
<td>.22&lt;sup&gt;a&lt;/sup&gt;</td>
<td>270 (54.3)</td>
<td>890 (60.2)</td>
<td>.02&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Link to resources, counseling, or crisis support</td>
<td>604 (59.6)</td>
<td>536 (55.0)</td>
<td>.04&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>276 (55.5)</td>
<td>864 (58.4)</td>
<td>.26&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Tools to focus on the positive events and influences in life</td>
<td>578 (57.1)</td>
<td>553 (56.8)</td>
<td>.90&lt;sup&gt;a&lt;/sup&gt;</td>
<td>267 (53.7)</td>
<td>864 (58.4)</td>
<td>.07&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Brain games to improve thinking</td>
<td>525 (51.8)</td>
<td>480 (49.3)</td>
<td>.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>257 (51.7)</td>
<td>748 (50.6)</td>
<td>.66&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>How to cope with COVID-19</td>
<td>406 (40.1)</td>
<td>409 (42.0)</td>
<td>.39&lt;sup&gt;a&lt;/sup&gt;</td>
<td>200 (40.2)</td>
<td>615 (41.6)</td>
<td>.60&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>A chatbot to help you with daily stress</td>
<td>352 (34.7)</td>
<td>293 (30.1)</td>
<td>.05&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>139 (28.0)</td>
<td>506 (34.2)</td>
<td>.01&lt;sup&gt;a,b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Chi-square test.
<sup>b</sup> P values <.05 and less than the Benjamini-Hochberg critical value were considered to be statistically significant.

**Employment Status**

The three most important DMHT components for essential workers and unemployed individuals were information or education (essential: mean 6.09, SD 2.70; unemployed: mean 6.09, SD 2.61); mindfulness/meditation (essential: mean 6.17, SD 2.55; unemployed: mean 5.94, SD 2.62); and link to resources, counseling, or crisis support (essential: 6.00, SD 2.89; unemployed: mean 5.86, SD 2.72). Unemployed participants were more likely to rate tools to focus on positive life events and influences as essential (mean 6.88, SD 2.62; unemployed: mean 6.66, SD 2.70); and a chatbot to help with daily stress (352/1013, 34.7%, vs 293/974, 30.1%; χ²=4.29, P=.04) as more important than essential workers.

When provided the option to build their own DMHT, the most common features listed by essential workers were information and education (618/974, 63.4%), mindfulness/meditation (584/974, 60.0%), and symptom tracking (tracking sleep or mood; 555/974, 57%). The most common features reported by unemployed persons was mindfulness/meditation (687/991, 67.8%), information or education (636/991, 62.8%), and distraction tools (eg, drawing, puzzles, music) (630/991, 62.2%). In comparing the desired features for a DMHT by employment status, unemployed participants were more likely to request that their DMHT include mindfulness/meditation (687/1013, 67.8% vs 584/974, 60.0%; χ²=13.31, P<.001); distraction tools (drawing, puzzles, and music; 630/1013, 62.2% vs 540/974, 55.4%; χ²=9.34, P=.002); link to resources, counseling, or crisis support (604/1013, 59.6%, vs 536/974, 55.0%; χ²=4.29, P=.04); and a chatbot to help with daily stress (352/1013, 34.7%, vs 293/974, 30.1%; χ²=4.93, P=.03) than the essential worker group (Table 5).

**Distress Status**

The most important DMHT components among distressed and nondistressed users included information or education (distressed: mean 6.01, SD 2.67; nondistressed: mean 6.32, SD 2.59); mindfulness/meditation (distressed: mean 6.09, SD 2.56; nondistressed: mean 5.96, SD 2.68); and link to resources, counseling, or crisis support (distressed: mean 5.95, SD 2.81; nondistressed: mean 5.88, SD 2.80). Distressed individuals also rated tools to focus on positive life events and influences as important (mean 5.90, SD 2.42).

When provided the option to build their own DMHT, nondistressed individuals indicated information or education (327/497, 65.8%), followed by mindfulness/meditation (305/497, 61.4%), distraction tools (276/497, 55.5%), and link to resources, counseling, or crisis support (276/497, 55.5%). Similarly, distressed individuals desired to include...
mindfulness/meditation (966/1479, 65.3%), followed by information or education (927/1479, 62.7%) and distraction tools (894/1479, 60.4%). Compared to nondistressed individuals, distressed participants preferred to include symptom tracking (270/497, 54.3% vs 890/1479, 60.2%; $\chi^2=5.25, P=.02$) and a chatbot (139/497, 28.0% vs 506/1479, 34.2%; $\chi^2=6.60, P=.01$) within their DMHT (Table 5).

### Discussion

#### Principal Findings

This study documented DMHT use among essential workers and unemployed individuals during the COVID-19 pandemic and determined which features such users would prefer to have in a DMHT offering. DMHT use has been deemed by many in the field to be subpar, and some have suggested that poor uptake and adherence to such tools is the result of user burden and inadequate match to user needs [17]. Indeed, our findings indicate that despite reports of increased downloads [48] and user registration by digital mental health companies [13], use of DMHTs by essential workers and those unemployed due to COVID-19 is very similar to prepandemic reports (14%).

When asked to build their own DMHT, individuals who did not use a DMHT to cope during the COVID-19 pandemic preferred to include information or education (1091/1680, 64.9%), mindfulness/meditation (1071/1680, 63.8%), and distraction tools (1031/1680, 61.4%). DMHT users preferred to include mindfulness/meditation (200/277, 72.2%), tools to focus on the positive events and influences in life (178/277, 64.3%), and symptom tracking (tracking sleep or mood; 166/277, 59.9%).

Participants who used DMHTs to cope during COVID-19 were more likely than those who did not use DMHTs to prefer mindfulness/meditation features (200/277, 72.2% vs 1071/1680, 63.8%; $\chi^2=7.46, P=.006$), positive psychology features (178/277, 64.3% vs 953/1680, 56.7%; $\chi^2=5.53, P=.02$), and chatbot features (108/277, 39.0% vs 537/1680, 32.0%; $\chi^2=5.31, P=.02$). Conversely, compared to non–DMHT users, DMHT users were less likely to prefer brain games to improve thinking (124/277, 44.8% vs 881/1680, 52.4%; $\chi^2=5.61, P=.02$), and distraction tools (139/277, 50.2% vs 1031/1680, 61.4%; $\chi^2=12.38, P<.001$) (Table 6).

### Table 6. Digital mental health tool (DMHT) features stratified by status.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Non–DMHT user (n=1680), n (%)</th>
<th>DMHT user (n=277), n (%)</th>
<th>P value</th>
<th>Total (N=1987), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mindfulness/meditation</td>
<td>1017 (63.8)</td>
<td>200 (72.2)</td>
<td>.006a,b</td>
<td>1271 (64.0)</td>
</tr>
<tr>
<td>Information or education</td>
<td>1019 (64.9)</td>
<td>163 (58.8)</td>
<td>.05a</td>
<td>1254 (63.1)</td>
</tr>
<tr>
<td>Distraction tools (drawing, puzzles, music)</td>
<td>1031 (61.4)</td>
<td>139 (50.2)</td>
<td>&lt;.001a,b</td>
<td>1170 (58.9)</td>
</tr>
<tr>
<td>Symptom tracking (tracking sleep or mood)</td>
<td>994 (59.2)</td>
<td>166 (59.9)</td>
<td>.81a</td>
<td>1160 (58.4)</td>
</tr>
<tr>
<td>Link to resources, counseling, or crisis support</td>
<td>986 (58.7)</td>
<td>154 (55.6)</td>
<td>.33a</td>
<td>1140 (57.4)</td>
</tr>
<tr>
<td>Tools to focus on the positive events and influences in life</td>
<td>953 (56.7)</td>
<td>178 (64.3)</td>
<td>.02a,b</td>
<td>1131 (56.9)</td>
</tr>
<tr>
<td>Brain games to improve thinking</td>
<td>881 (52.4)</td>
<td>124 (44.8)</td>
<td>.02a,b</td>
<td>1005 (50.6)</td>
</tr>
<tr>
<td>How to cope with COVID-19</td>
<td>689 (41.0)</td>
<td>126 (45.5)</td>
<td>.16a</td>
<td>815 (41.0)</td>
</tr>
<tr>
<td>A chatbot to help you with daily stress</td>
<td>537 (32.0)</td>
<td>108 (39.0)</td>
<td>.02a,b</td>
<td>645 (32.5)</td>
</tr>
</tbody>
</table>

*a*Chi-square test.

bP values <.05 and less than the Benjamini-Hochberg critical value were considered to be statistically significant.
emotional, physical, financial, and privacy burdens were seen as acceptable, with essential workers finding these tools to be more burdensome than the unemployed group. Increased perceived burdensome may be partially explained by previous findings suggesting that essential workers have increased fatigue from elevated anxiety and work demands during the ongoing pandemic [3].

Individuals with increased mental health needs (ie, the distressed group) reported more financial burden of DMHTs than the nondistressed. It is understandable that during a pandemic, where people are struggling financially, there would be concerns about the costs of DMHTs, given that many popular and widely publicized tools require a paid subscription. In the United States, those who lost their jobs during COVID-19 are faced with insufficient insurance to cover the costs of mental health care [51], and those who are struggling financially likely have additional financial concerns aside from a DMHT subscription fee, such as the cost of data plans and the technology needed to use these services. In fact, an earlier study noted that most individuals with depression and/or anxiety symptoms preferred using health apps that were free or had low cost for download (eg, <$5) [43]. As such, reimbursement is one part of the solution for increasing access to care for everyone, and until technology is more affordably available to all, the use of these services will be compromised [52].

When asked to design their own DMHT for coping with COVID-19, again mindfulness/meditation was listed as an important feature for all subgroups in this study. Interestingly, information and education about COVID-19 was also consistently listed as an important feature in all subgroups except for people who had used DMHTs during the pandemic. In addition to mindfulness/meditation, people who used DMHTs to cope with COVID-19 preferred positive psychology tools and mood and sleep tracking. Figure 2 illustrates the preferences between the unemployed and essential worker groups. This finding has important implications for DMHT development focused on pandemic response and other prolonged environmental disasters. Developers would be able to create a single tool that includes mindfulness/meditation, information and education about COVID-19 coping, and distraction tools, which would appeal to a wide group of people with different needs during COVID-19, with only a few added features for specific populations.

A final finding in this study was reasons for not turning to DMHTs to cope with COVID-19. Most of the sample indicated that they did not use a DMHT because they did not think to look for such a tool. Past reports suggest that this result may be due to a lack of information about how DMHTs might be effective [53]. This assumption is further supported by the fact that one-third of the sample did not think a DMHT would be helpful to them, and one-quarter of the sample indicated that they had other means of coping. The potential lack of confidence in DMHTs might be addressed through education to health providers on the effectiveness of DMHTs [54], the creation of reimbursement codes in the United States that would allow providers to prescribe these services [55], or the further use of a human-centered design from DMHT companies to create tools that are appropriately targeting user needs and concerns.

Comparison With Prior Work

A strength of this study is that we explicitly asked a large sample of users about their app preferences and perceived importance of various features. This survey was different from previous studies that have primarily focused on downloads and user metrics [48], insight from providers and private digital health companies [56], and self-report from individuals exclusively with mild depression or mild anxiety symptoms with exclusion of severer mental health conditions (eg, suicidality) [43]. It is also novel in its consideration of user-centered design principles (eg, ease of use and learnability) when developing and identifying DMHT features that would be most acceptable to a very large sample of potential target consumers. Consistent with emerging models that integrate community-based research, implementation science, and user-centered design principles [57,58], this is an important first step in a well-planned process of DMHT design to identify the needs and preferred features that users, both experienced and unexperienced, and preferences for what tools they would like to see in a DMHT. Previous studies that used self-report of physical health and mental health apps found that users typically only use an app for one feature [43]. It might be that future apps need to have multiple features incorporated to meet the overarching needs of similar populations. As Mohr et al [17] have noted previously, health app developers tend to create a tool based on what the developer feels is essential and historically only designs around these developer-driven features, rather than asking the end-user what role they see digital health playing in their lives, what needs they have that are unmet, and what functions they want these tools to have. By starting with understanding end-user needs and preferences, DMHT developers may see not only an increase in DMHT uptake but long-term use as well.

The findings of this study differ from findings in recent studies on the use of technology to cope with the consequences of COVID-19. According to recent research in the general population, there has been increased desire for apps or online resources that allow for fitness at home, owing to physical distancing and stay-at-home orders that have led to a shift from gyms and group fitness classes to exercise at home [59]. During pre-pandemic times, Rubanovich et al [43] found that people with depression and anxiety symptoms reported more frequently using health apps featuring fitness, pedometers, or heart rate monitoring apps than DMHTs. Conversely, in our study, fitness apps and tools were listed very low in the list of tools participants used for coping with COVID-19. Although studies on the use of fitness apps among essential workers and employee groups are sparse, existing research suggests that the use of such tools in practice is low [60], which may explain why these tools were not in the top group of DMHTs listed by these participants. According to past research, those who are unemployed may likewise not have resources to engage in fitness apps, and generally are less likely to engage in fitness tracking [61]. Finally, another COVID-19 study found that more contact tracing and COVID-19 informational apps were being downloaded than DMHTs in North America [62]. We note here that downloads are often not equivalent to tool use as recent research has found that many people do download such tools but rarely use them long term [20,63]. Our study specifically
asked about which DMHTs people used to cope with COVID-19 stress.

Our study adds to the existing body of work by understanding how DMHTs could be made to be more accessible to those at risk for the emotional consequences of COVID-19. Many experts in digital mental health have argued for the need to better personalize such tools [54] and to include the perspectives of the intended consumer in the design of such tools [8].

Limitations
Although this study has important implications regarding the use of DMHTs from a human-centered design approach, it does have limitations. First, this is a cross-sectional study surveying the US participants’ experiences and opinions at one point in time. Second, the participants of this sample are likely to be more accepting of digital tools, as they were recruited from an online research platform. As such, the information from this study is limited to those who are currently using and familiar with technology. Third, this study did not consider cross-cultural acceptance of DMHTs, which is an important caveat since a DMHT may be different in countries that already support such tools as part of their health care system. Fourth, we are unable to explicitly comment on the sample’s overall experience with apps or DMHTs during prepandemic times. The focus of this paper was to explore whether users were using available, low-cost DMHTs to address COVID-19–related stress. Future studies should conduct a more thorough assessment of both current and previous DMHT use.

Conclusions
Despite the limitations, this study provides important information to the mental health care system and to those who develop and provide DMHTs during prolonged stressful events. Policy makers and providers may not be able to rely on existing DMHTs to address the emotional health of essential workers and people who are unemployed. This study points to the need to ensure DMHTs address the needs that the intended consumer feels is most important, that these tools are not burdensome under high-stress conditions, and that they are affordable to people who have limited means.

Acknowledgments
The research reported in this publication was supported by the National Institutes of Health (grant P50MH115837 and T32MH020021). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions
PA, KAC, and FM-G contributed to study concepts and design. PA and KAC obtained funding. MJ and FM-G conducted or interpreted the statistical analyses. MP consulted on the analytic approach. FM-G, MJ, and PA drafted the manuscript with contributions from all the authors. All authors read and approved the final manuscript for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Consent form and COVID-19 mental health apps survey.
[DOC File, 92 KB - mental_v8i8e28360_app1.doc]

Multimedia Appendix 2
Distress measures stratified by app users.
[DOC File, 74 KB - mental_v8i8e28360_app2.doc]

Multimedia Appendix 3
Additional supplementary tables.
[DOCX File, 53 KB - mental_v8i8e28360_app3.docx]

References


39. Help me find an app. OneMindPsyberguide. URL: https://onemindpsyberguide.org/apps/ [accessed 2021-07-05]

40. Explore relevant apps and reviews. M-Health Index and Navigation Database. URL: https://apps.digitalpsych.org/ [accessed 2021-07-05]


Abbreviations

ANOVA: analysis of variance
CAGE-AID: Cut-Annoyed-Guilty-Eye Adapted to Include Drugs
DMHT: digital mental health tool
GAD-2: 2-item Generalized Anxiety Disorder
MIND: M-Health Index and Navigation Database
PHQ-2: 2-item Patient Health Questionnaire
SBQ-R: Suicide Behaviors Questionnaire–Revised
UBS: Use Burden Scale
SARS: severe acute respiratory syndrome
SUS: System Usability Scale
VOIP: voice over IP