JMIR Mental Health

Journal Impact Factor (JIF) (2023): 4.8
Volume 6 (2019), Issue 5   ISSN 2368-7959   Editor in Chief: John Torous, MD, MBI

Contents

Original Papers

Viewpoint

Original Paper

An Approach for Data Mining of Electronic Health Record Data for Suicide Risk Management: Database Analysis for Clinical Decision Support

Sofian Berrouiguet[1,2,3,4], MD, PhD; Romain Billot[4], PhD; Mark Erik Larsen[5], PhD; Jorge Lopez-Castroman[6], MD, PhD; Isabelle Jaussent[7], MD, PhD; Michel Walter[1], MD, PhD; Philippe Lenca[4], PhD; Enrique Baca-García[8,9,10,11,12,13,14], MD, PhD; Philippe Courtet[7,10,11,15,16], MD, PhD

[1]Adult Psychiatry, Brest Medical University Hospital at Bohars, Brest, France
[2]EA 7479 Soins Primaires, Santé Publique, Registre des Cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France
[3]Mental Health Department, University Hospital of Brest, Brest, France
[4]F-29238, Laboratoire des Sciences et Techniques de l'information de la Communication et de la Connaissance, IMT Atlantique, Brest, France
[5]Black Dog Institute, University of New South Wales, Sydney, Australia
[6]Inserm U1061, Crisis Admission Center, University Hospital of Nîmes, Nîmes, France
[7]Inserm U1061, La Colombières Hospital, University of Montpellier, Montpellier, France
[8]Carlos III Institute Of Health, Centro de Investigation en Salud Mental, Madrid, Spain
[9]Department of Psychiatry, Universitad Catolica del Maule, Talca, Chile
[10]Department of Psychiatry, General Hospital of Villaba, Madrid, Spain
[11]Department of Psychiatry, University Hospital Rey Juan Carlos, Mostoles, Spain
[12]Department of Psychiatry, University Hospital Infanta Elena, Valdemoro, Spain
[13]Psychiatry Department, Universidad Autónoma de Madrid, Madrid, Spain
[14]Department of Psychiatry, Instituto de Investigación Sanitaria de la Fundación Jiménez Díaz, Madrid, Spain
[15]Department of Emergency Psychiatry and Acute Care, Centre Hospitalier Universitaire de Montpellier, Université Montpellier, Montpellier, France
[16]Fondamental Foundation, Créteil, France

Corresponding Author:
Sofian Berrouiguet, MD, PhD
Adult Psychiatry
Brest Medical University Hospital at Bohars
Route de Ploudalmezeau
Brest, 29200
France
Phone: 33 668204178
Email: sofian.berrouiguet@gmail.com

Abstract

Background: In an electronic health context, combining traditional structured clinical assessment methods and routine electronic health–based data capture may be a reliable method to build a dynamic clinical decision-support system (CDSS) for suicide prevention.

Objective: The aim of this study was to describe the data mining module of a Web-based CDSS and to identify suicide repetition risk in a sample of suicide attempters.

Methods: We analyzed a database of 2802 suicide attempters. Clustering methods were used to identify groups of similar patients, and regression trees were applied to estimate the number of suicide attempts among these patients.

Results: We identified 3 groups of patients using clustering methods. In addition, relevant risk factors explaining the number of suicide attempts were highlighted by regression trees.

Conclusions: Data mining techniques can help to identify different groups of patients at risk of suicide reattempt. The findings of this study can be combined with Web-based and smartphone-based data to improve dynamic decision making for clinicians.

(JMIR Ment Health 2019;6(5):e9766) doi:10.2196/mental.9766

KEYWORDS

clinical decision support system; data mining; electronic health; mobile phone; prevention; suicide; suicide attempts

Introduction

Suicide Risk Assessment

Over 800,000 people die of suicide every year, and it is estimated that for each suicide, there may have been >20 other attempted suicides. A previous attempt is the major predictor of death by suicide [1]. However, many other outcomes associated with suicidal behaviors should be considered in the preventive and therapeutic decision-making process for effective prevention [2]. Thus, clinical [3], environmental [4], and genetic [2] suicide risk factors have been intensively studied among suicide attempters. Indeed, attempters provide data to identify suicide-related risk factors, and such at-risk patients are a privileged target for proper prevention and intervention strategies (eg, by mitigating risk factors or by maintaining contact with clinical support) [5]. Empirically informed suicide risk assessment frameworks are useful in guiding the evaluation and treatment of individuals presenting with suicidal symptoms. Actual guidelines recommend the systematic identification of risk factors based on risk assessment scales [6]. Nevertheless, the limits of actual risk assessment procedures may be a false reassurance for clinicians, and the conflation of risk assessment and risk prediction may be confusing to clinicians [7]. Therefore, there is an urgent need for an innovative tool that could integrate both empirical and structured assessment to support decision making in suicide prevention.

Clinical Decision-Support Systems

Decision-support tools help providers in their decision-making process. The use of these tools has been on the rise in recent years owing to their ability to bring evidence-based medicine to the point of care. A clinical decision-support system (CDSS) is a health information system that is integrated into electronic health records (EHR), enabling easy and effective use by physicians [8]. The CDSS incorporates individual patient data, a rule engine, and a medical knowledge base to produce a patient-specific assessment or recommendation of a management plan [9]. The CDSS usually relies on the processing of clinical data gathered into EHRs. However, these techniques have been poorly explored in mental health and the suicide-prevention setting [10].

Thus, there is still an important need to develop a CDSS that supports clinician decision makers to choose, for example, the most appropriate treatment, the nature of a psychosocial strategy, or the duration of treatment in suicide prevention strategies. A key feature of such a CDSS is to identify a patient’s risk in terms of a repeated attempt, the number of reattempts, or suicide death within a period of time. The development of both passive and active collection of patients’ data provides the opportunity to improve clinician knowledge and thus determine risk factors and relevant combinations of risk factors [11].

Aims

This study aims to combine data from EHRs to provide support to decision making for clinicians in suicide prevention. We present the main results of a data mining process on a sample of suicide attempters to first identify groups of similar patients and then identify risk factors associated with the number of suicide attempts. We hypothesize that a data mining process helps to better characterize the population of suicide attempters by identifying the most relevant groups of patients and their associated risk factors for suicide reattempt (or other variables of interest). The ultimate goal is to build a CDSS for clinician decision support and propose a personalized prevention and intervention strategy to each patient.

Methods

Patient Recruitment

Suicide attempters aged >18 years were recruited from consecutive admissions to the Emergency Department or specialized Acute Care Unit of three university hospitals (University Hospital Ramon y Cajal, Madrid, Spain; Fundación Jimenez Diaz, Madrid, Spain; and Academic Hospital of Montpellier, Montpellier, France) between 1994 and 2006. Owing to their specific characteristics [12,13], major suicide repeaters were excluded. One of the hospitals is part of the Spanish National Health System and the other, the French National Health System; both hospitals provide medical coverage for all emergencies in a catchment area covering a population of around 500,000 people in Madrid and 400,000 in Montpellier. After providing a complete description of the study to participants, written informed consent was obtained. Trained psychiatrists or psychologists interviewed all patients before discharge. The study was approved by the local research ethics committees in Madrid and Montpellier (CPP Montpellier Sud-Méditerranée IV, CHU Montpellier). The research followed the Code of Ethics of the World Medical Association (Declaration of Helsinki). Protocols and assessment procedures in both centers are based on the Columbia Suicide History Form [14].

Procedure and Clinical Assessment

The French or Spanish version of the Mini-International Neuropsychiatric Interview (MINI) [15] was used to obtain Axis I Diagnostic and Statistical Manual of Mental Disorders - 4th edition diagnoses. Psychiatric diagnoses were classified in the following categories: mood disorder (specifying depression or bipolar disorder), anxiety disorders, obsessive-compulsive disorder, alcohol or drug misuse, psychotic disorders, eating disorders, somatoform disorders, and adjustment disorders. The lifetime diagnosis was determined using a best-estimate procedure. The psychiatrist in charge of the patient’s care assigned the diagnosis based on MINI interviews, medical records, and information from relatives, when available.

Suicide risk was assessed using the Suicide Intent Scale [16], a semistructured 15-item rating scale yielding a global score that indicates the severity of suicidal intent. The Risk-Rescue Rating Scale [17] is a 10-item interviewer-administered scale designed to assess the lethality and intent of a suicide attempt, measuring the life risk derived from it, and the likelihood of a rescue intervention at the time of the attempt.

Statistical Methods

A robust data-qualification process was performed to ensure data quality and consistency before statistical analyses. Although data were intended to be collected according to the same clinical procedures, quality variations were expected between the hospitals. Missing data were identified, and a variable was retained only when the completion rate reached 70%. When needed, new variables were created. For instance, 34 answers of the Barratt Impulsiveness Scale survey, version 10 (BIS10), were treated to build 3 scores of impulsiveness in terms of motor impulsivity, attentional impulsivity, or nonplanning impulsivity [18]. For each question of the BIS10, the score ranged from 1 (low impulsivity) to 4 (high impulsivity). The total score ranges from 34 to 136 points. The subscores ranged from 11 to 44 points for motor and attentional impulsivity and from 12 to 48 points for nonplanning impulsivity.

Unidimensional and two-dimensional analyses for both quantitative and qualitative variables were carried out. In addition, Fisher-Snedecor procedures were used to compare the two subgroups (male vs female) when needed. An unsupervised approach was used to extract homogeneous patterns from the data without any prior hypothesis. The approach is based on a multiple correspondence analysis (MCA) of qualitative variables to reduce the dimensionality. It consists of representing patients in a factorial space where each dimension is a combination of initial variables. Quantitative variables (eg, age) are not used during the calculations but are projected onto the factorial space. Hierarchical Clustering on Principal Components is then performed from the patients' representation in the initial factorial space. Hierarchical clustering has many advantages, including the construction of a hierarchical tree called dendrogram that enables a visual interpretation of the dataset. The dendrogram depicts the emergence of groups of patients who share common risk patterns. In addition, it facilitates discussion between statisticians and practitioners to choose the optimal number of clusters. Each cluster was then interpreted through the association between the cluster and the list of qualitative and quantitative variables (V test). In the second step, the focus was on the variable of interest—the number of suicide attempts. Recursive partitioning has been used as a multivariable procedure that classifies individuals (patients) by successively splitting into subpopulations. Furthermore, a regression tree was built, and the number of suicide attempts was explained by different binary tests on predictive variables.

Results

Population Description

From the original database, the first step relied on data qualification. Several redundancies (eg, duplicated surveys or alternative coding) were observed among 263 initial variables. Subsequently, a completion threshold was applied to the resulting variables, and only 23 variables satisfied a 70% minimum completion rate. Three additional variables related to the types of impulsivity (as described above) were added. With respect to the 2802 initial patients, we decided to keep only suicide attempters with a 100% completion rate for the 26 variables. In the final filtering, 5 variables were disregarded for redundancy or useless purpose (the type of patients, assessment date, source, and year and day of birth). This rigorous process ensured high data quality for both patients and variables; it also provided a final dataset of 681 patients and 21 variables. Participants were predominantly young (mean age 40.1 years), female, employed, and married. Most patients included in the final analysis had a history of mental disorders, including major depression (482/681, 70.8%), bipolar disorder (160/681, 23.0%), dysthymic disorder (30/681, 4.4%), obsessive-compulsive disorder (58/681, 8.5%), and alcohol misuse (178/681, 26.1%) (Table 1).

Table 1. Clinicosociological main features of the postfiltering dataset of 681 suicide attempters.

Features	Value
General features	
Qualitative variables, n (%)	
Sex	
Female	493 (72.4)
Male	188 (27.6)
Marital status	
Single	239 (35.1)
Married	240 (35.2)
Separated or divorced	181 (26.6)
Widowed	21 (3.1)
Children	
No	272 (39.9)
Yes	409 (60.1)
Education	
Low	31 (4.6)
Intermediate	368 (54.0)
High	282 (41.4)
Employment	
Employed	451 (66.2)
Unemployed	110 (16.2)
Incapacity	41 (6.0)
Retired	79 (11.6)
Quantitative variable, age (years), median (Q1-Q3)	40.6 (28-49.6)
Clinical features	
Qualitative variables, n (%)	
History of mental disorder	
No	6 (0.9)
Yes	675 (99.1)
History of family suicidal behavior	
No	424 (62.3)
Yes	257 (37.7)
Lifetime major depression	
No	199 (29.2)
Yes	482 (70.8)
Lifetime bipolar disorder	
No	521 (76.5)
Yes	160 (23.5)
Lifetime dysthymic disorder	
No	651 (95.6)
Yes	30 (4.4)
Lifetime obsessive-compulsive disorder	
No	623 (91.5)

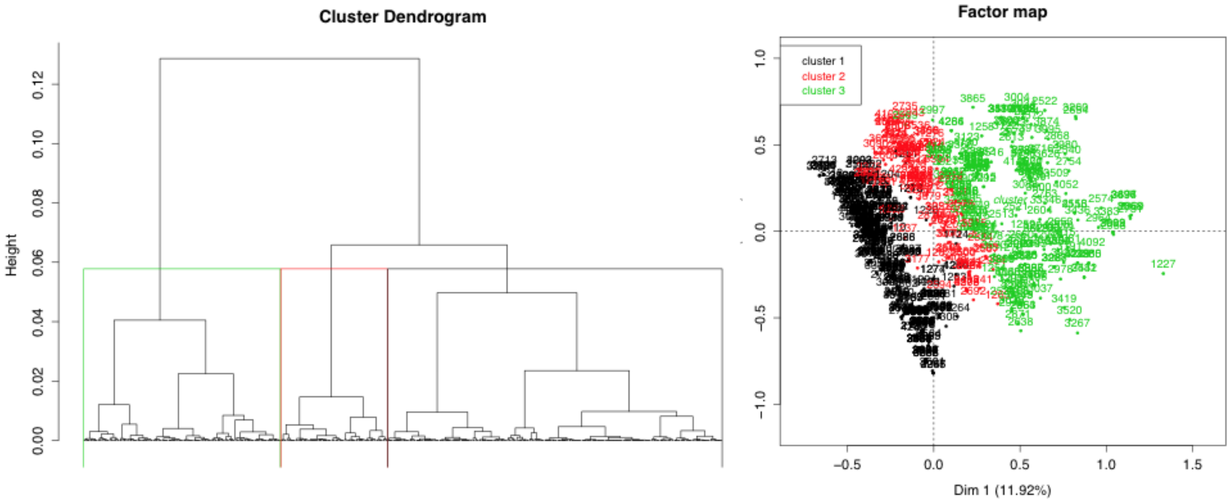
Features	Value
Yes	58 (8.5)
Lifetime eating disorder	
No	571 (83.8)
Yes	110 (16.2)
Lifetime alcohol-drug misuse	
No	465 (68.3)
Yes	216 (31.7)
Lifetime substance misuse	
No	586 (86.0)
Yes	95 (14.0)
Lifetime alcohol abuse	
No	503 (73.9)
Yes	178 (26.1)
Quantitative variable	
Number of suicide attempts, median (Q1-Q3)	2 (1-3)
Barratt Impulsiveness Scale survey - version 10, scores (range)	
Motor impulsivity	26 (22-30)
Attentional impulsivity	27 (23-30)
Nonplanning impulsivity	28 (24-31)

Principal Outcome: Clustering of Patients

The first step of the analysis was to perform an MCA to reduce the dimension, followed by a hierarchical clustering from the principal components to highlight groups of homogeneous patients. The tree structure (in terms of inertia gain) and a discussion between statisticians and practitioners allowed us to study patients partitioned into three clusters. Figure 1 shows the cluster dendrogram and the projection of the three clusters onto the factor map; the factor map represents a two-dimensional projection of the first two dimensions only.

We conducted an in-depth analysis of the 3 groups for data interpretation. Statistical association tests (V tests) enabled identification of over- or underrepresented modalities in the three clusters. Cluster 1 was mainly related to an average patient profile of women (positive association V test, *P*<.001) who did not misuse drugs, substances, or alcohol (*P*<.001) and without bipolar disorder (*P*<.001), but with previous or current episodes of depression (*P*<.001) and other mental health disorders (*P*=.01). Cluster 3 was, in contrast, associated with men (*P*<.001) and drug, substance, and alcohol misuse (*P*<.001); in this group, patients were mainly single (*P*=.01), with no children (*P*=.006) and no experience of depression (*P*=.006). Compared with clusters 1 and 3, cluster 2 was neutral in terms of gender, but this group was related to people with a work incapacity (*P*=.03), low education level (*P*=.02), possible bipolar disorder (*P*<.001), and no drugs or alcohol misuse (*P*<.001) and without episodes of depression or other mental health disorders (*P*<.001). Thus, without any prior hypothesis, this unsupervised approach underlined three homogeneous groups. Gender appeared as a crucial marker for two of the three groups.

Figure 1. Hierarchical clustering (left) and multiple correspondence analysis factor map (right) with three projected clusters.

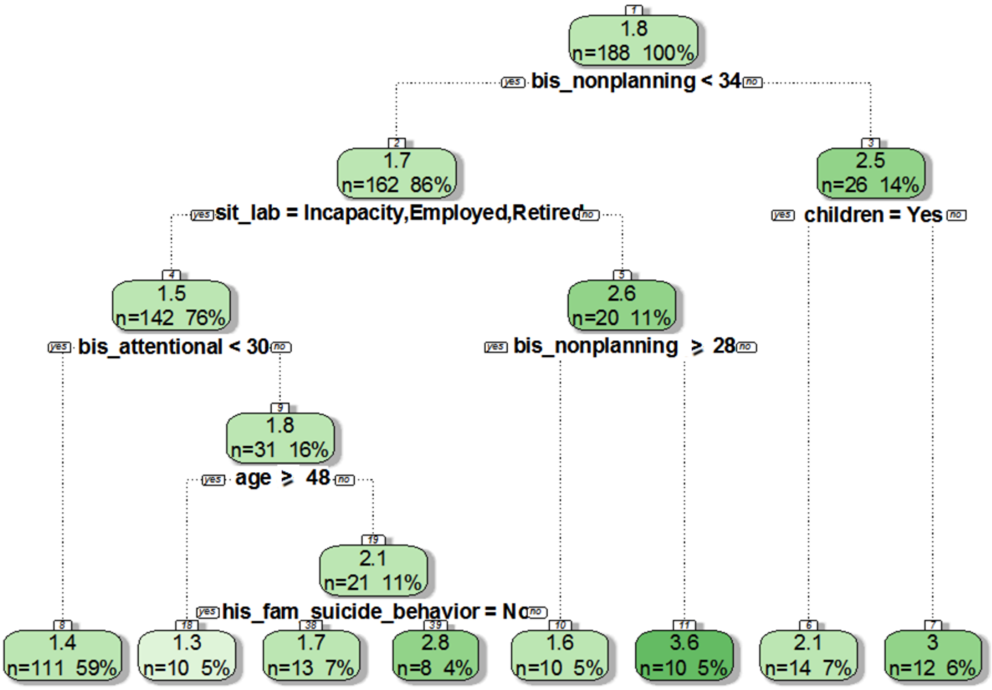


Secondary Outcome: Identification of Factors Associated With the Risk of Repeated Suicide Attempts

The second step of the analysis aimed to identify factors associated with a higher risk of suicide attempts (variable “number of suicide attempts”) for men and women separately, following the principal outcome. Figure 2 depicts the regression tree for male patients, while Figure 3 shows the results for female patients only. For both groups, we noted that impulsivity aspects (including the motor section, nonplanning section, and attentional impulsivity section of the BIS interview) were relevant factors explaining the number of suicide attempts for a patient. Furthermore, higher scores were associated with a higher number of attempts for patients.

While analyzing both groups, the first conclusion is a clear difference between genders. For instance, eating disorders are linked to a higher number of suicide attempts for women (mean 2.9 in women vs 2.3 in men, *P*=.005), while a history of familial suicidal behavior (mean 2.8 in men vs 1.7 in women, *P*<.001) and the employment status are risk factors for some men. In particular, unemployed men with higher scores at the BIS interview were at higher risk of suicide repetition. Older age and having children were also identified as risks factors.

Figure 2. The decision tree on the variable "number of suicide attempts" according to gender "male". BIS: Barratt Impulsiveness Scale.



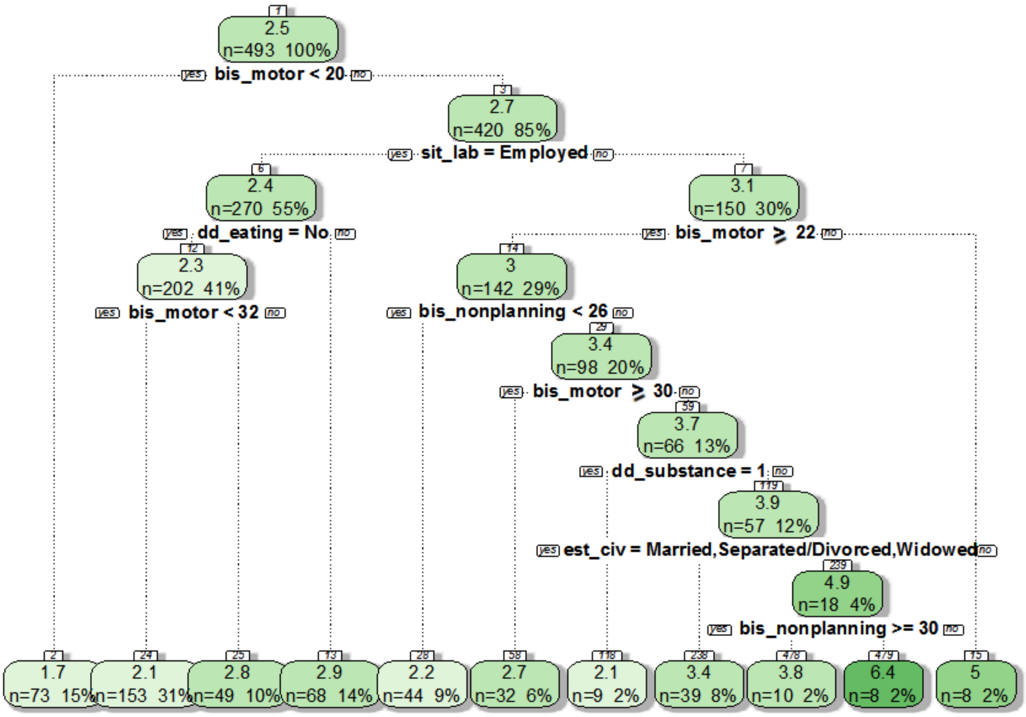


Figure 3. The decision tree on the variable “number of suicide attempts” according to gender “female”. BIS: Barratt Impulsiveness Scale.

Discussion

Principal Findings and Integration into a Dynamic Clinical Decision-Support System

A systematic assessment before discharge from hospital has allowed building of a large database suitable for modern data mining techniques. In this study, we identified clusters of suicide attempters and variables that may explain the repetition of suicide attempt in suicide attempters. This study shows how a simple structuration of the assessment of discharged patients after a suicide attempt may provide relevant data for clustering methods. The clustering may help clinicians allocate a patient into a risk cluster. Therefore, it is the first step of the CDSS design. This model may lead to a stratified approach in decision making for suicide prevention. Furthermore, analyzing larger datasets could allow the discovery of new risk factors that are not currently considered relevant during clinical interviews. However, we did not propose a model for suicide prediction; our model mining big databases is a prerequisite toward better decision making for suicide prevention. Furthermore, this model could also be applied to other data sources like personal health records or ecological momentary assessment (EMA).

Our findings are in line with recent studies showing how suicide risk assessment could lead to patient clustering from a preventative perspective [19]. Eleven clinically relevant items related to the characteristics of suicidal behavior were submitted to a Hierarchical Ascendant Classification; the results showed that most individuals were included in a cluster characterized by less lethal means and planning (“impulse-ambivalent”). The second cluster featured more carefully planned attempts (“well-planned”), more alcohol or drug use before the attempt, and more precautions to avoid interruptions. Finally, the third cluster included individuals reporting more attempts (“frequent”), more often serious or violent attempts, and an earlier age at the first attempt. In addition, differences across clusters by demographic and clinical characteristics were found, particularly with the third cluster whose participants had experienced high levels of childhood abuse. Overall, a systematic, structured assessment may help clinicians characterize suicide risk better and personalize prevention strategies. We believe that electronic health and data mining techniques may help us to reach this goal.

In this study, participants were assessed by trained clinicians before discharge from the ED. Data were captured using paper-based formularies of the actual MeMind Web-based EHR [20]; these data could have been captured by our MeMind Web app [20] designed to gather observational data through an EHR interface and perform EMA [18,19]. This Web-based software has two distinct views—the EHR view (for clinicians) and the EMA view (for patients). The “EHR” view is designed to be used by doctors and nurses during the face-to-face assessment. As current EHRs, the app also collects sociodemographic, diagnostic, and pharmacological treatment information within the “assessing suicide” protocol. Sociodemographic variables included age (defined as the age at the index episode), sex, profession, current working status, marital status, number of children (if any), and educational level. Moreover, family history of suicidal behavior, age at the first suicide attempt, and violence of the suicide attempt were measured. The EMA view for patients to track their symptoms was not used in this study. Clinicians and patients can access the Web app either from a computer or their personal mobile phone. EMA involves repeated sampling of subjects’ behaviors and experiences in real time in their natural environment. EMA has been successfully used for real-time self-reporting of symptoms and behavior. For example, Husky et al showed the utility and feasibility of using EMA to study suicidal ideation [4]. For this study, patients did not have any access to this interface. Such a Web app may represent the future of suicide risk assessment, as it allows data mining of static data (EHR data provided by a single clinical assessment) and dynamic data (smartphone data provided by EMA techniques); these data can be processed to build a dynamic CDSS [19].

Limitations

Focus on Suicide Attempters

In this study, patients were recruited after a suicide attempt. We postulate that the development of a CDSS would be more relevant in a population of suicide attempters. Suicide attempters are also defined as an “indicated population” [2] who warrant the maximal attention of health care service owing to the risk of reattempt. Treatment and follow-up strategies are well described in the guidelines for suicide attempters. Therefore, we were able to propose treatment to each participant based on actual recommendations. However, applying these strategies to samples of patients with suicide ideation but without a history of suicide attempt would be hypothetical and ethically controversial. Owing to the epidemiological specificities of both populations, a specific model would have to be built for each population. In terms of multiple suicide attempters, a relevant perspective is to focus on the temporality of suicide attempts rather than the absolute number; to achieve this objective, finer data are needed.

Missing Data: The Data Mining Challenge

This study illustrates the need for high-quality and large databases for extracting significant patient profiles or risk factors. In this study, starting from an initial set of 2802 patients with 263 variables, the data-qualification process resulted in a final sample of 681 patients with 21 complete variables. Although this volume of data already ensures statistical significance, it underlines the importance of better ways to standardize data collection in participating institutions. The CDSS quality strongly depends on input data. Moreover, a critical challenge may be the clinician acceptance of such tools that directly impact the completion rate of the EHR [21]. Another option could be the integration of other data sources, such as personal health records and EMA [21]. Overall, the use of larger databases will refine different profiles of patients and dynamically improve the personalized prevention strategies, thanks to the EMA data.

Recommendations for Suicide Prevention

From Guidelines to Clinical Decision-Support Systems in Suicide Prevention

Guidelines recommend that all patients presenting to the hospital services with self-harm should receive a psychological assessment before discharge, to determine the risk of further reattempt [6]. This assessment should also help clinicians choose the most appropriate treatment while considering clinical guidelines and patient-specific risks factor. However, reviews have addressed the challenge clinicians working in the emergency setting face when they rely on these tools to perform decision making [7]. As EHRs are extensively used in emergency services and psychiatric departments, we propose integrating CDSS features regarding an individual’s risks factors into EHRs.

Toward Dynamic Clinical Decision-Support Systems

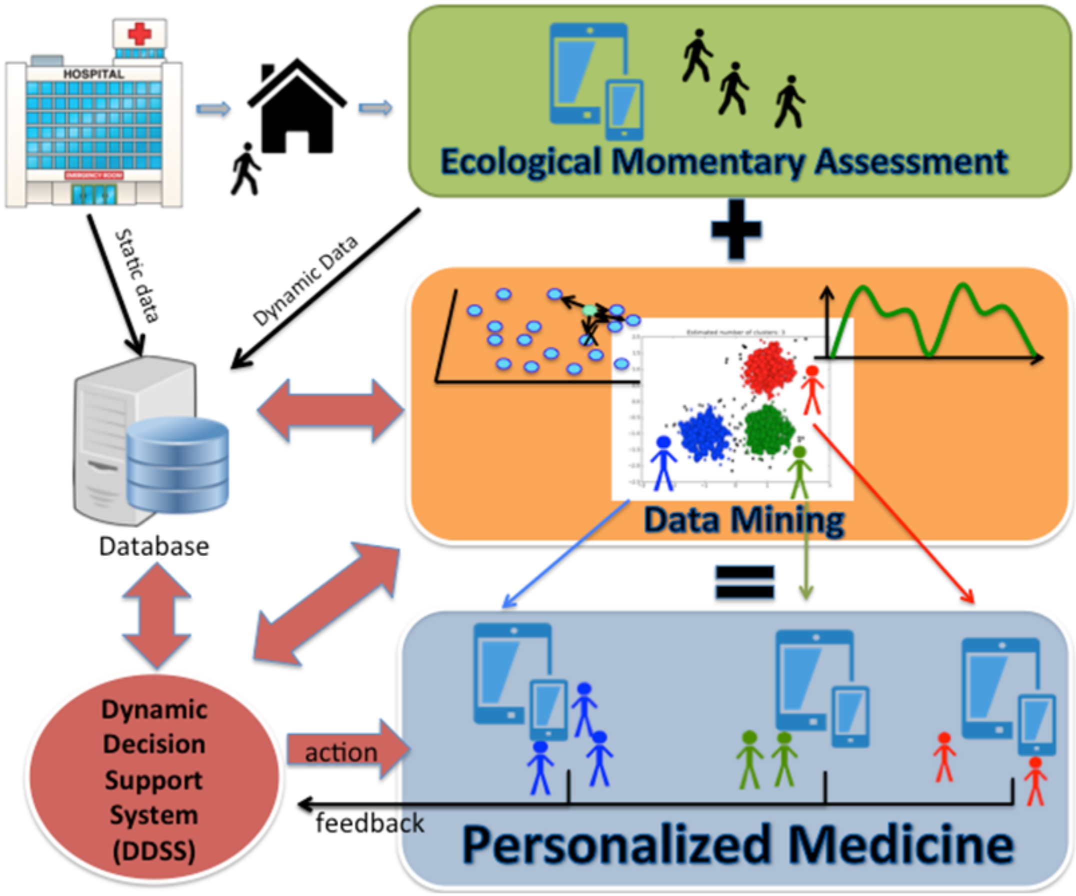
Most clinicians have use EHRs daily in emergency services and psychiatric units. However, few institutions have taken advantage of recent technological advances opportunities in risk assessment. Combining electronic health–based assessment with data mining techniques represents an opportunity to foster suicide-prevention research. This new paradigm is useful in providing personalized intervention strategies by itself, but it also affords the opportunity to identify novel mechanisms to be targeted in suicide-prevention strategies. In addition, we believe that computational models can provide data-assisted ideas emerging from these repositories and will have special appeal for the empirically minded clinicians [7].

Although studies have highlighted the value of self-reports in clinical assessment, they are rarely routinely implemented [22]. Internet and mobile technologies are ideal for self-monitoring assessment and ecological observational studies. Mobile phones are generally kept on at all times and carried everywhere, making them an ideal platform for the broad implementation of EMA technology. For example, Husky et al conducted a study providing support for the use of EMA in the identification of suicidal ideation in outpatients [23]. These techniques may help clinicians identify risky events occurring during follow-up. Overall, these EMA data could also be mined and integrated into the decision-making process.

Conclusions

The next step is to take advantage of new technologies and current developments of Web-based mobile apps to design the next-generation dynamic CDSS (Figure 4). As emerging mobile health (mHealth) techniques in suicide prevention strategies also produce relevant data, this study proposes a new model of the decision-support system based on the data mining proceedings from face-to-face assessment and mHealth EMA. Our strategy relies on the processing of static data (initial assessment) and dynamic data (EMA) able to instantaneously deliver to clinicians decision support regarding a specific patient. Indeed, mHealth apps allow patients to report their physical or mental health conditions and symptoms, hence providing dynamic data able to enrich studies, confirming or rejecting statistical hypotheses [24]. Such dynamic data based on EMA will form the core of a dynamic decision-support system, which will adapt its recommendations to patients’ characteristics (Figure 4).

Figure 4. The decision-support system based on ecological momentary assessment and data mining.



Acknowledgments

Authors acknowledge the French *Groupement d’Etude et de Prevention du Suicide* (GEPS) and the French Embassy in Madrid, Spain. This study received a Hospital Clinical Research Grant (PHRC 2009) from the French Health Ministry. None of the funding sources had any involvement in the study design; collection, analysis, or interpretation of data; writing of the report; or the decision to submit the paper for publication. This study was funded partially by Instituto de Salud Carlos III (ISCIII PI13/02200; PI16/01852), Delegación del Gobierno para el Plan Nacional de Drogas (20151073), and the American Foundation for Suicide Prevention (LSRG-1-005-16).

Conflicts of Interest

None declared.

References

1. Finkelstein Y, Macdonald EM, Hollands S, Sivilotti MLA, Hutson JR, Mamdani MM, Canadian Drug SafetyEffectiveness Research Network (CDSERN). Risk of Suicide Following Deliberate Self-poisoning. JAMA Psychiatry 2015 Jun;72(6):570-575. [doi: 10.1001/jamapsychiatry.2014.3188] [Medline: 25830811]
2. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016 Jul;3(7):646-659. [doi: 10.1016/S2215-0366(16)30030-X] [Medline: 27289303]
3. Chiu HY, Lee HC, Chen PY, Lai YF, Tu YK. Associations between sleep duration and suicidality in adolescents: A systematic review and dose-response meta-analysis. Sleep Med Rev 2018 Dec;42:119-126. [doi: 10.1016/j.smrv.2018.07.003] [Medline: 30093362]
4. Husky M, Swendsen J, Ionita A, Jaussent I, Genty C, Courtet P. Predictors of daily life suicidal ideation in adults recently discharged after a serious suicide attempt: A pilot study. Psychiatry Res 2017 Dec;256:79-84. [doi: 10.1016/j.psychres.2017.06.035] [Medline: 28624676]

5. Berrouiguet S, Courtet P, Larsen ME, Walter M, Vaiva G. Suicide prevention: Towards integrative, innovative and individualized brief contact interventions. Eur Psychiatry 2018 Jan;47:25-26. [doi: 10.1016/j.eurpsy.2017.09.006] [Medline: 29096129]
6. Chan MKY, Bhatti H, Meader N, Stockton S, Evans J, O'Connor RC, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. Br J Psychiatry 2016 Dec;209(4):277-283. [doi: 10.1192/bjp.bp.115.170050] [Medline: 27340111]
7. Mulder R, Newton-Howes G, Coid J. The futility of risk prediction in psychiatry. Br J Psychiatry 2016 Dec;209(4):271-272. [doi: 10.1192/bjp.bp.116.184960] [Medline: 27698212]
8. Kruse CS, Mileski M, Vijaykumar AG, Viswanathan SV, Suskandla U, Chidambaram Y. Impact of Electronic Health Records on Long-Term Care Facilities: Systematic Review. JMIR Med Inform 2017 Sep 29;5(3):e35 [FREE Full text] [doi: 10.2196/medinform.7958] [Medline: 28963091]
9. Celi LA, Zimolzak AJ, Stone DJ. Dynamic clinical data mining: search engine-based decision support. JMIR Med Inform 2014 Jun 23;2(1):e13 [FREE Full text] [doi: 10.2196/medinform.3110] [Medline: 25600664]
10. Karmakar C, Luo W, Tran T, Berk M, Venkatesh S. Predicting Risk of Suicide Attempt Using History of Physical Illnesses From Electronic Medical Records. JMIR Ment Health 2016 Jul 11;3(3):e19 [FREE Full text] [doi: 10.2196/mental.5475] [Medline: 27400764]
11. Glenn C, Nock M. Improving the short-term prediction of suicidal behavior. Am J Prev Med 2014 Sep;47(3 Suppl 2):S176-S180 [FREE Full text] [doi: 10.1016/j.amepre.2014.06.004] [Medline: 25145736]
12. Mendez-Bustos P, de Leon-Martinez V, Miret M, Baca-Garcia E, Lopez-Castroman J. Suicide reattempters: a systematic review. Harv Rev Psychiatry 2013;21(6):281-295. [doi: 10.1097/HRP.0000000000000001] [Medline: 24201820]
13. Blasco-Fontecilla H, Jaussent I, Olié E, Béziat S, Guillaume S, Artieda-Urrutia P, et al. A cross-sectional study of major repeaters: a distinct phenotype of suicidal behavior. Prim Care Companion CNS Disord 2014;16(4) [FREE Full text] [doi: 10.4088/PCC.14m01633] [Medline: 25664212]
14. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
15. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34 [FREE Full text] [Medline: 9881538]
16. Beck R, Morris J, Beck A. Cross-validation of the Suicidal Intent Scale. Psychol Rep 1974 Apr;34(2):445-446. [doi: 10.2466/pr0.1974.34.2.445] [Medline: 4820501]
17. Weisman A, Worden J. Risk-rescue rating in suicide assessment. Arch Gen Psychiatry 1972 Jun;26(6):553-560. [Medline: 5027119]
18. Patton J, Stanford M, Barratt E. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995 Nov;51(6):768-774. [Medline: 8778124]
19. Berrouiguet S, Barrigón ML, Brandt SA, Nitzburg GC, Ovejero S, Alvarez-Garcia R, et al. Ecological Assessment of Clinicians' Antipsychotic Prescription Habits in Psychiatric Inpatients: A Novel Web- and Mobile Phone-Based Prototype for a Dynamic Clinical Decision Support System. J Med Internet Res 2017 Jan 26;19(1):e25 [FREE Full text] [doi: 10.2196/jmir.5954] [Medline: 28126703]
20. Memind. URL: https://www.memind.net/es/app/security/login/ [accessed 2019-02-28] [WebCite Cache ID 76X1AbIJZ]
21. Beglaryan M, Petrosyan V, Bunker E. Development of a tripolar model of technology acceptance: Hospital-based physicians' perspective on EHR. Int J Med Inform 2017 Dec;102:50-61. [doi: 10.1016/j.ijmedinf.2017.02.013] [Medline: 28495348]
22. Davidson CL, Anestis MD, Gutierrez PM. Ecological Momentary Assessment is a Neglected Methodology in Suicidology. Arch Suicide Res 2017 Jan 02;21(1):1-11. [doi: 10.1080/13811118.2015.1004482] [Medline: 26821811]
23. Husky M, Olié E, Guillaume S, Genty C, Swendsen J, Courtet P. Feasibility and validity of ecological momentary assessment in the investigation of suicide risk. Psychiatry Res 2014 Dec 15;220(1-2):564-570. [doi: 10.1016/j.psychres.2014.08.019] [Medline: 25155939]
24. Larsen ME, Nicholas J, Christensen H. A Systematic Assessment of Smartphone Tools for Suicide Prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]

Abbreviations

CDSS: clinical decision-support system
EHR: electronic health records
EMA: ecological momentary assessment
MCA: multiple correspondence analysis
mHealth: mobile health
MINI: Mini-International Neuropsychiatric Interview

Edited by G Wadley, R Calvo, M Czerwinski, J Torous; submitted 10.02.18; peer-reviewed by P Winkler, K Krysinska, A Kotlo; comments to author 08.03.18; revised version received 30.06.18; accepted 26.07.18; published 07.05.19.

Please cite as:
Berrouiguet S, Billot R, Larsen ME, Lopez-Castroman J, Jaussent I, Walter M, Lenca P, Baca-García E, Courtet P
An Approach for Data Mining of Electronic Health Record Data for Suicide Risk Management: Database Analysis for Clinical Decision Support
JMIR Ment Health 2019;6(5):e9766
URL: https://mental.jmir.org/2019/5/e9766/
doi:10.2196/mental.9766
PMID:31066693

Original Paper

A Group-Based Online Intervention to Prevent Postpartum Depression (Sunnyside): Feasibility Randomized Controlled Trial

Jennifer Duffecy[1], PhD; Rebecca Grekin[2,3], PhD; Hannah Hinkel[2,4], BA; Nicholas Gallivan[2,5], BA; Graham Nelson[2], MA; Michael W O'Hara[2], PhD

[1]Department of Psychiatry, College of Medicine, University of Illinois at Chicago, Chicago, IL, United States

[2]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, United States

[3]Serious Mental Illness Treatment Resource and Evaluation Center, Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States

[4]Western Carolina University, Cullowhee, NC, United States

[5]Kansas State University, Manhattan, KS, United States

Corresponding Author:
Jennifer Duffecy, PhD
Department of Psychiatry
College of Medicine
University of Illinois at Chicago
912 S Wood
Chicago, IL, 60612
United States
Phone: 1 3124131225
Email: jduffecy@gmail.com

Abstract

Background: Postpartum depression (PPD) has a 20% 3-month prevalence rate. The consequences of PPD are significant for the mother, infant, and the family. There is a need for preventive interventions for PPD, which are effective and accessible; however, many barriers exist for women who attempt to access perinatal depression prevention programs. Internet interventions for the treatment and prevention of depression are widely accepted as efficacious and may overcome some of the issues pertaining to access to treatment barriers perinatal women face. However, internet interventions offered without any human support tend to have low adherence but positive outcomes for those who do complete treatment. Internet support groups often have high levels of adherence but minimal data supporting efficacy as a treatment for depression. Taken together, these findings suggest that combining the treatment components of individual interventions with the support provided by an internet support group might create an intervention with the scalability and cost-effectiveness of an individual intervention and the better outcomes typically found in supported interventions.

Objectives: This study aimed to describe the development of a cognitive behavioral therapy (CBT) internet intervention with peer support to prevent PPD and examine preliminary depression and site usage outcomes.

Methods: User-centered design groups were used to develop the internet intervention. Once the intervention was developed, women who were 20 to 28 weeks pregnant with symptoms of depression (Patient Health Questionnaire-8 scores of 5-14) but who had no major depression diagnosis were enrolled in a randomized controlled trial (RCT) to compare 8 weeks of a CBT-based internet intervention with peer support to an individual internet intervention designed to prevent PPD. Assessments took place at baseline, 4 weeks, 8 weeks (end of treatment), and then 4 weeks and 6 weeks postpartum.

Results: A total of 25 women were randomized. Of these, 24 women completed the RCT. Patient Health Questionnaire-9 scores at 6 weeks postpartum remained below the clinical threshold for referral for treatment in both groups, with depression measures showing a decrease in symptoms from baseline to postpartum. At 6 weeks postpartum, only 4% (1/24) met the criteria for PPD. There was no difference between groups in adherence to the intervention, with an average of 14.55 log-ins over the course of treatment.

Conclusions: Results suggest women were responsive to both peer support and individual internet interventions to prevent PPD and that peer support may be a useful feature to keep participants adherent.

Trial Registration: ClinicalTrials.gov NCT02121015; https://clinicaltrials.gov/ct2/show/NCT02121015 (archived by WebCite at http://www.webcitation.org/765a7qBKy)

(JMIR Ment Health 2019;6(5):e10778) doi:10.2196/10778

KEYWORDS

postpartum depression; cognitive behavioral therapy; internet; social support

Introduction

Background

Postpartum depression (PPD) has a 20% 3-month prevalence rate [1]. The consequences of PPD are significant for the mother, infant, and the family [2]. Loss of pleasure; low mood; fatigue; difficulty thinking, concentrating, and making decisions; and sleep and appetite disturbance all lead to impairment in daily functioning, especially in caring for an infant. Depressed mothers often show gaze avoidance, more negative and fewer positive facial expressions, and slower or mistimed responses to infant bids for attention [3]. Infants of depressed mothers show less eye gaze during feeding, less play, less positive affect, and higher levels of withdrawal behavior; are more drowsy and fussy; and show higher levels of insecure attachment than infants of nondepressed mothers [2,4]. This negative impact extends to the preschool years and beyond [4,5]. In sum, PPD leads to impairment in maternal behavior, cognition, and affect, which has a clear and negative long-lasting effect on the child. The need for preventive interventions for PPD, which are effective and accessible, is widely recognized.

Many types of face-to-face PPD prevention interventions have been developed and tested. Cognitive behavioral therapy (CBT), interpersonal psychotherapy, and antidepressant medications have demonstrated benefit in preventing the onset of depressive symptoms [6,7]. CBT has considerable support in preventing major depressive episodes as well perinatal depression. CBT focuses on teaching patients to identify and challenge dysfunctional beliefs and problematic behaviors. It has been shown to be effective in numerous settings, including individual-, group-, and distance-based interventions [8]. However, many barriers exist for women who attempt to access traditional perinatal depression prevention or treatment programs. Barriers may exist on multiple levels (patient, provider, and practice) [9] and make it difficult for women to access services, even if they have been identified as having depressive symptoms or being at risk for developing them. Barriers include stigma, cost, scheduling difficulties, and lack of providers and programs [10-12]. Structural barriers such as time constraints, lack of child care, and transportation also are substantial barriers to obtaining psychological treatment among 75% of depressed urban primary care patients [13]. Online prevention and intervention programs have the potential to overcome many of these barriers.

Internet interventions for the treatment of depression are widely accepted as efficacious [14-18]. Though prevention interventions are fewer, they have demonstrated support [19-21]. These interventions commonly consist of didactic material and interactive tools to practice skills, such as cognitive restructuring or behavioral activation. Stand-alone individual interventions (those without any human support) typically show smaller effect sizes than coach-supported interventions, likely because of decreased adherence. Although coach-supported interventions may have improved outcomes, drawbacks include increased cost and decreased scalability. Internet support groups (ISGs), where peers provide support rather than trained coaches, are frequently utilized by pregnant women and new mothers, with 75% of mothers endorsing use of internet-delivered support [22]. However, despite good adherence and high levels of interest, ISGs have limited data supporting their efficacy for treating or preventing depression [23-26].

Objectives

Taken together, these findings suggest that combining the treatment components of individual interventions with the support provided by ISGs may create an intervention with the scalability and cost-effectiveness of an individual intervention and the outcomes similar to those found in coach-supported interventions. There have been a number of small trials that suggest that peer support has the potential to improve adherence and depressive outcomes [23,27,28], and this study sought to examine the use of this approach in the perinatal population. The aims of this study are to describe the development of a CBT internet intervention with peer support to prevent PPD and examine preliminary depression and site usage outcomes.

Methods

Stage I: User Centered Design Groups

User-Centered Design Groups

With institutional review board approval, user-centered design groups were conducted to engage women from the target population in the intervention-building process (topics, site motif, and usability of potential application). User-centered design groups were held from July to September 2014, with a total of 6 participants across 3 groups. A total of 2 groups were held in-person on campus at 1 of the participating institutions; the third group was held online for ease of data collection and participation. All were facilitated and analyzed by a psychologist consultant for this grant and analyzed using a phenomenological approach to qualitative analysis [29]. No software was used.

Participants and Procedures

Participants were recruited by word-of-mouth from investigators at the 2 participating institutions and through recruitment flyers distributed around the campus of 1 institution. Participants ranged in age from 25 to 45 years, reported median income of US $100,000.00 and education level as graduate or postgraduate. Furthermore, 100% of the sample identified as white, and 100% were partnered or married. Of them, 1 participant was still pregnant (0 children), 2 participants reported having 2 children each, and 3 participants reported 1 child in their household.

Before beginning the groups, a moderator guide was developed to ensure consistency of questions and topics covered. Low-fidelity prototypes were used to obtain feedback on various site components and motifs. All groups were shown the same set of materials. Questions were formulated along 3 lines of inquiry: (1) pregnancy topics of interest (topics about which women might be seeking more information); (2) motifs for the intervention (visual themes and look and feel of the internet site); and (3) use of the intervention (how, when, and why they might interact with the intervention).

Investigators defined pregnancy topics of interest in initial meetings regarding the creation of the intervention and vetted these with group participants. The initial list was developed by consulting the literature and brainstorming among the authors and collaborators, who have extensive clinical experience with pregnant women. Investigators also asked for any topics participants did not see included in the premade list. Participants were initially asked for motifs they had either seen or thought of themselves in response to pregnancy, mothering, babies, etc. In total, 3 motifs for the intervention had already been created by the research team in conjunction with the application development team and were vetted among group participants. Finally, participant interest in using the proposed intervention was gauged using diagrams of predetermined themes, potential topics as discussed in the focus group, and the broader concept of social media for information and community. Predetermined motifs were developed through brainstorming sessions with the authors, program designers, and graphic designers.

Stage II: Pilot Trial

Study Design

Participants were randomized in groups of 7 to 9 to either the Share (group) condition or the Control (individual) condition. In total, the Share condition comprised 18 women, and the Control condition comprised 7 women. Randomization of 2:1 was utilized to gain more experience with the group intervention component, which was more novel than the individual intervention. Contact was lost with 1 participant in the Share condition immediately after randomization, so 17 women actually participated in the Share condition. Subsequent to randomization, participants had an initial 20-min phone call with the study staff to establish rapport, ensure site functionality, and elicit change-talk via motivational interviewing. Study staff had no additional contact with participants except for regularly scheduled assessments. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Iowa (UI) [30].

Website

The Sunnyside website was an 8-week online prevention intervention developed by research partners at the Northwestern University [31]. The Sunnyside website is based on cognitive behavioral principles [32] that consisted of 16 core didactic lessons (plus 3 postpartum booster sessions) and 5 associated tools. See Multimedia Appendix 1 for an overview of intervention content. CBT was selected as the intervention approach given the evidence base supporting its effectiveness in managing perinatal depression as well as the authors' experience in creating online CBT interventions [31,33]. Each lesson, which required 10-15 min to complete, was uniquely designed to provide information about pregnancy and postpartum issues, as well as the components of CBT. At the conclusion of each lesson, women were prompted with a *Call to Action* slide that encouraged them to directly apply the CBT strategies that were learned in the lessons. The lessons comprised text and video material. In addition to the 8 weeks that comprised the core portion of the intervention, participants also completed postpartum booster sessions, which were lessons made available at 2 weeks, 4 weeks, and 6 weeks postpartum.

New lessons and accompanying tools were released twice a week, and participants received an email notification upon the release of this material. The initial lesson introduced the cognitive behavioral principles utilized throughout the intervention and explained how one's thoughts and behaviors affect their moods and physical being. The program contained 5 interactive cognitive behavioral tools: thought restructuring (*Think*), mood tracking (*Feel*), activity scheduling and monitoring (*Do*), relaxation (*Relax*), and goal setting (*Achieve*). Associated tools served to complement the lessons by having the women directly apply the CBT strategies that were discussed in the lessons. See Multimedia Appendix 1 for more information on the content of the lessons and tools. Taken together, the lessons and tools provided useful information and additional resources on how to manage mood and cope with depression and anxiety.

Experimental Intervention: Share

Content and general layout was identical for the Control and Share, with the exception that the Share site featured a newsfeed and accountability features. Although the Control participants focused on the lessons and interactive tools on the website, the Share participants also collaborated through the *Activity Feed*. The *Activity Feed* was a constantly updating feed that displayed each of the women's activity on the site. Here, participants were able to post, *like*, and comment or provide feedback to other women's posts. Discussion questions were posted with the release of each lesson to encourage interaction.

Share participants also maintained an *Individual Garden Plot* as well as a *Community Garden* that were linked to user profiles. Women provided information about themselves in the profiles to increase group bond. In both of the garden plots, incentives, such as garden gnomes or flower collections, were earned by completing various tasks on the site, such as reading a new lesson or posting on the feed, but they were only added to the Community Garden once each of the group members had completed the identified task. The flower garden provided a visual representation of each participant's site use to increase accountability to each other on task completion. Women were also able to reach out to each other with a generic *nudge* message that sent an email indicating that a specific group member requested return to the site. A deliberate decision was made to not allow private messaging between participants to encourage interaction on the site.

Women were also provided a *Contact Moderator* tool to report any issues that arose with either the site or with group members. Staff members watched over the site to verify that medically inaccurate information was not being posted and to ensure this was a safe space for participants to disclose their feelings.

Measures

Participants completed a total of 5 study assessments, which included self-reported online questionnaires and interview-based assessments that were conducted over the phone by trained graduate students in clinical psychology. Outcomes were assessed at baseline, week 4 of the treatment program, week 8 (end of treatment), 4 weeks postpartum, and 6 weeks postpartum. Interview assessments were conducted at baseline, week 8, and 6 weeks postpartum. Self-report measures were collected at all 5 time points. Assessors were blinded to site arm and intervention usage. Participants were compensated US $20 per assessment completed, for a total of US $100 over the course of the study. Compensation was tied to assessment completion, not site usage, to decrease loss to follow-up in assessment even if participants were no longer engaged with the site.

Use

Use was examined through total number of log-ins, completion of tools, and lessons. Peer support features (likes, comments, nudges, and posts) were examined for those in the Share condition.

Usability and Satisfaction

Usability and satisfaction were measured using the 17-item Usability, Satisfaction and Ease of Use questionnaire (USE) [34] which was designed to measure satisfaction, usefulness, ease of use, and ease of learning on a 1-7 Likert scale, with higher numbers indicating greater usability and satisfaction.

Depressive Symptoms

The Hamilton Depression Rating Scale (HDRS) is a 17-item scale that assesses the severity of depression symptoms. Participants are scored on a range of severity (0-4) or incidence (0-2), based on the variable. Variables include depressed mood, agitation, and somatic symptoms, among others [35]. Data on the structured HDRS support interrater reliability, internal consistency, and high test-retest reliability [36].

The Inventory of Depression and Anxiety Symptoms (IDAS) is a self-report tool that aims to assess specific dimensions of depression and anxiety symptomatology. It contains 10 symptom-specific scales, including suicidality, appetite, and panic, among others, as well as 2 broader scales for general depression and dysphoria. This inventory possesses both internal consistency and content validity [37]. The 20-item general depression scale was used for this study.

The Patient Health Questionnaire (PHQ-8) is an 8-item modification of the Primary Care Evaluation of Mental Disorders used to provide diagnostic criteria for depression symptoms and commonly used in depression screening. The Patient Health Questionnaire-9 (PHQ-9) is a replica of the 8-item form with the addition of a suicidality item. Participants are scored based on the frequency of certain moods and behaviors over the last 2 weeks, from 0 (not at all) to 3 (nearly every day) [38]. The internal reliability and test-retest reliability of this measure was excellent with a Cronbach alpha of .89 and .84, respectively, and correlates strongly with other mental health assessments [39]. The PHQ-8 was administered during the initial online screening, whereas the PHQ-9 was used for all other assessments.

Psychiatric Diagnosis

The Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders (DSM) Axis-I Disorders (SCID-I) is a semistructured interview that guides the diagnosis of the major DSM Axis-I disorders. Its modules show excellent to good reliability and superior validity compared with clinical interviews (First, 1995). Suicidality was assessed using the suicide question from the Mini International Neuropsychiatric Interview (MINI), a brief structured interview for the diagnosis of DSM and International Classification of Diseases disorders [40].

Statistical Analyses

As this was a pilot study, descriptive statistics were used to examine the data. Results should be interpreted with caution because of the small sample size.

Procedures

Eligibility

Eligible participants were 18 years of age or older, between 20- and 28-weeks gestation at the time of baseline assessment, had a score between 5 and 14 on the PHQ-8 screener (mild-moderate depressive symptoms), were able to read and speak English, and had access to the internet on any device. Exclusion criteria included diagnosis of a major depressive episode, psychotic disorder, bipolar disorder, substance use disorder or other diagnoses using the MINI, current use of psychotropic medications, intention to resume antidepressant medication after delivery (if women discontinued use during pregnancy), currently in psychotherapy, and endorsed suicidality (with separate procedures in place for responding to these women). The rationale for the PHQ-8 eligibility criteria is that subthreshold depression symptoms are a common criterion for entry to prevention programs [41,42]. The purpose of prevention interventions is to halt the progress toward PPD; thus, identifying those already on the trajectory is important.

Recruitment

All women who met eligibility criteria from May 2015 to August 2015 were invited to participate in the study. Participants were identified by the UI's Institute for Clinical and Translational Science (ICTS). The ICTS accessed the medical records from the UI Hospitals and Clinics and generated a list of women who were currently pregnant and 18 years of age and older. UI research team members followed up with the women generated from that list to inquire about interest in participation. In addition, UI and University of Illinois-Chicago employed use of mass-email and advertisement via Research Match.

Interested participants were directed to the study Web page where they completed the initial online screener. Consent was obtained before completion of the screening questionnaire. Eligible women were invited to complete the baseline clinical phone interview and self-report assessments, which were the final step in determining eligibility for the study. Once women completed the assessments and were deemed eligible, they were randomized into a cohort. This is discussed in greater detail below.

All procedures were approved by the institutional review boards at the UI, Northwestern University, and University of Illinois-Chicago.

Results

Stage I: User Centered Design Groups

Topics

Topics to be included in the final version of the intervention fell within 3 common categories—physical changes (mother and fetus), logistical challenges, and emotional stresses.

In terms of physical changes, 50% (3/6) of women reported that including information on body changes—such as weight, body after baby, changes during and after pregnancy, and breastfeeding—would be helpful. Participants suggested that simple psychoeducation itself would ease anxiety levels but they had difficulty finding useful information. Logistical challenges—including the broad *different strategies work for different kids* and the specific *When to amend your will* were the second most common responses in terms of topic category. Finally, 100% (6/6) of women described some sort of emotional challenge they had anticipated at some time during their pregnancy, including the fears of *Am I meant to be a mother* and *the experience of having a baby is not “easy” or “happy” or “blissful*.” Each woman described including these topics in an app designed to prevent PPD or at least recognize it early on if it were to occur, as a helpful, if not essential, goal.

Motifs

In reviewing predetermined themes offered for the overall appearance of the app, the overwhelming favorite (6/6, 100%) was a flower garden. This theme seemed to reflect the undercurrent of *new* and *emerging* in spontaneous participant answers. In total, 1 participant stated:

> *...the flower theme...I think it has a positive tone of growing and new life.*

Participants also felt strongly and negatively toward the predetermined themes of Egg in a Nest and Fish Bowl. Of them, 1 participant commented:

> *Not sure about the eggs...reminds me of something to eat.*

The same participant more strongly added:

> *I don’t know what I would want—but I know I wouldn’t want anything with teddy bears, ducks, or other baby animals.*

Use

Participants were asked a number of questions about their anticipated or hypothetical use of an intervention during pregnancy. Participants suggested notifying program participants of new content via any number of messaging systems, including text, email, push notification, or pop-up within the app. Participants suggested the information in the intervention should be credible, monitored by an authority or third-party figure, and should include information about how to use the app, as well as how to be social on the app. In this way, study participants could create a supportive, informative community. Finally, participants suggested incentivizing social interaction on the intervention with changes in the flower garden environment, such as growth or more flowers in the garden as interpersonal activity or use of the app increases.

Stage II: Pilot Trial

Participants

A total of 24 pregnant women (1 additional woman ceased participation immediately after randomization) in their second trimester participated in exchange for compensation. The sample was predominantly white (17/24, 72%), with 8% (2/24) African American, 8% (2/24) Asian or Asian American, 8% (2/24) multiracial, and 4% (1/24) Latina. The average age was 30.5 years (SD 4.05). The majority of the sample was married or cohabitating (20/24, 83%) and were employed either part-time or full-time (16/24, 67%). PHQ-8 scores at screening were 6.7 (SD 1.6). In total, 216 women completed the online screener. Most women screened out during the online screening process (n=180). Of those who completed the baseline assessment, 83% (*n*=25) were eligible and enrolled in the trial, which is a similar rate to other online depression prevention trials [21]. A total of 5 women (17%) were excluded because of current major depressive disorder (MDD) symptoms. See Figure 1 for more information.

Attrition and Site Use

Overall, 1 participant in the Share group withdrew from participation before beginning the intervention. Mean number of log-ins across the 8-week intervention plus booster sessions was 12.6 (SD 6.9) for Control (N= 7) and 16.5 (SD 9.3) for Share (N= 17). The average number of lessons accessed (Control: mean 12.1, SD 5.6; Share: mean 13.7, SD 4.3) and tools utilized (Control: mean 31.0, SD 28.6; Share: mean 24.3, SD 17.6) was similar between groups. Data are provided in Table 1.

Use of peer support features was fairly low (mean 5.9, SD 5.6); however, 59% (10/17) of the participants used at least one feature. Commenting on discussion questions or posts was most popular (59%); initiating status updates also was common (53%). The *like* and *nudge* features were utilized by fewer participants (35% and 12%, respectively).

Figure 1. CONSORT flow diagram. MDD: major depressive disorder; PHQ-8: Patient Health Questionnaire-8.

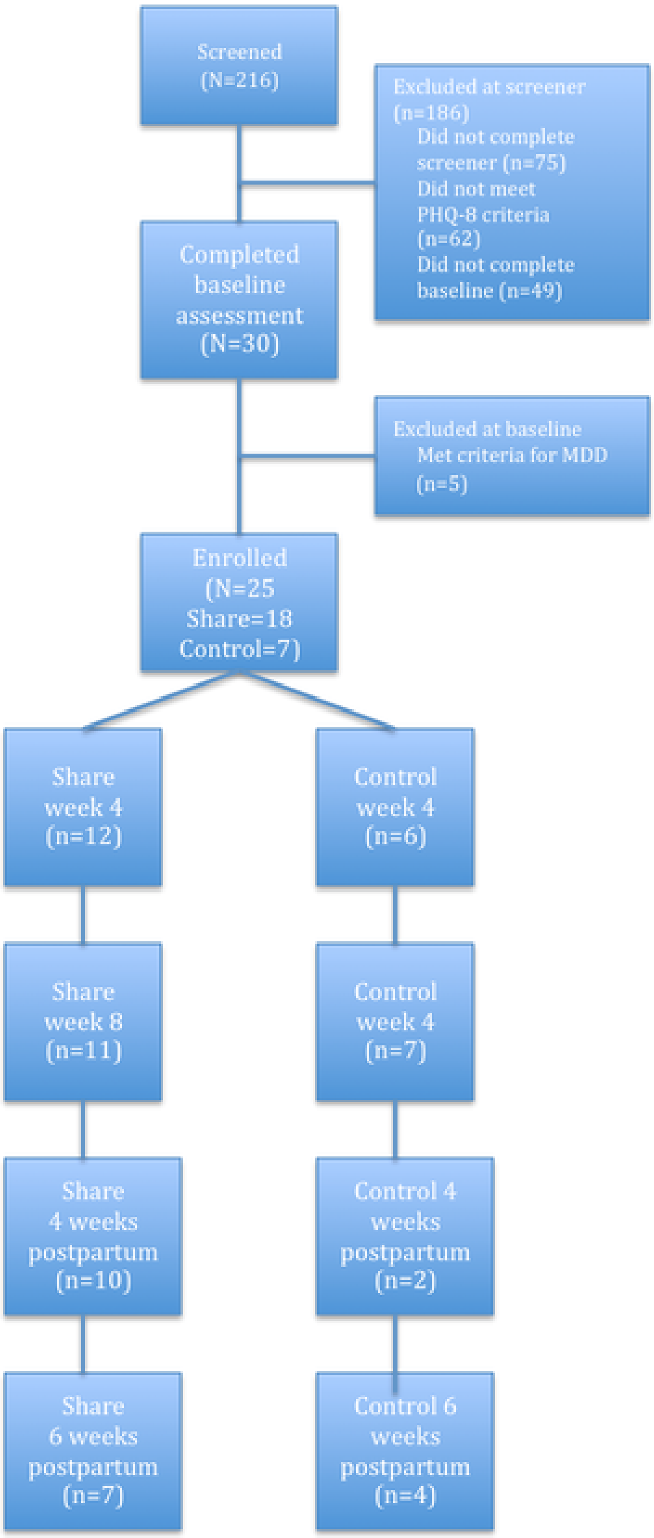


Table 1. Adherence data.

Program activity	Control (N=7)		Share (N=7)	
	Mean (SD)	Range	Mean (SD)	Range
Log-ins	12.6 (6.9)	5-21	16.5 (9.3)	2-35
Lessons accessed	12.1 (5.6)	2-17	13.7 (4.3)	2-18
Tool: activity scheduling/monitoring	10.7 (15.6)	1-34	6.2 (4.1)	1-12
Tool: mood rating	15.8 (9.9)	5-30	12.4 (9.5)	1-33
Tool: relaxation	7.7 (2.5)	5-10	3.4 (1.8)	1-6
Tool: thought record	6.5 (3.8)	1-11	5.1 (2.8)	1-11
Tool: goal setting	3.0 (0.0)	3-3	2.7 (3.1)	1-11
Total tools used	31.0 (28.6)	1-81	24.3 (17.6)	5-63

Usability and Acceptability

At week 8, scores on USE subscales ranged from 1 (strongly disagree) to 7 (strongly agree) and fell in the average range for both groups. Control participants' mean scores were 3.69 (SD 1.76) for usefulness, 4.94 (SD 0.88) for ease of use, 5.77 (SD 0.91) for ease of learning, and 3.93 (SD 1.70) for satisfaction. Share participants' mean scores were 4.42 (SD 0.87) for usefulness, 4.97 (SD 0.74) for ease of use, 5.73 (SD 0.87) for ease of learning, and 4.58 (SD 1.11) for satisfaction.

Depression

PHQ-9 scores at 6 weeks postpartum remained below 10, the clinical threshold for referral for treatment in both groups [39], with measures showing a decrease in symptoms from baseline to postpartum with the exception of the PHQ-9. PHQ-9 scores in the Control group increased slightly from baseline to 6 weeks postpartum, but still remained below a clinically significant threshold. Of them, 1 woman in the Control group met the criteria for a major depressive episode at 6 weeks postpartum diagnosed via the SCID. See Table 2 for more information.

Table 2. Depression outcomes.

Outcome measures	Baseline	4 weeks	8 weeks	4 weeks postpartum	6 weeks postpartum	Difference
Control, mean (SD)						
PHQ-9[a]	5.6 (3.5)	4.4 (3.7)	7.3 (4.8)	3.0 (0)	7.2 (5.4)	+1.6 (1.9)
HDRS[b]	8.3 (2.9)	—[c]	6.2 (3.9)	—	5.5 (4.0)	−2.8 (1.1)
IDAS[d]	47.3 (7.8)	42.4 (9.8)	41.6 (13.7)	45.0 (7.1)	41.3 (9.0)	−6.0 (1.2)
SCID[e]	0	—	0	—	1	—
Share, mean (SD)						
PHQ-9	5.1 (3.6)	5.8 (4.3)	3.7 (2.3)	2.3 (1.3)	3.7 (3.8)	−1.4 (.2)
HDRS	8.6 (5.4)	—	3.6 (2.0)	—	3.3 (2.3)	−5.3 (3.1)
IDAS	45.4 (6.3)	43.9 (6.6)	37.7 (4.7)	36.9 (7.0)	36.6 (4.8)	−8.8 (1.5)
SCID	0	—	0	—	0	—

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

[b] HDRS: Hamilton Depression Rating Scale.

[c] Not applicable.

[d] IDAS: Inventory of Depression and Anxiety Symptoms.

[e] SCID: Structured Clinical Interview for DSM Disorders.

Discussion

Principal Findings

This study outlines the development and initial testing of a novel Web-based intervention to prevent PPD. It indicates that pregnant women are willing to use an individual intervention or a group-based program and that doing so may impact the development of depressive symptoms. At the completion of the trial, only 1 woman (5%) in this at-risk sample met criteria for PPD compared with a 17% prevalence rate seen in at-risk women in the absence of an intervention [43]. Depressive symptoms decreased across most measures from baseline to postpartum in both the Share group and the Control individual condition. Given the adverse effects of depression on both mothers and babies, the reduction in symptoms is positive.

Intervention use was high as compared with other online PPD prevention interventions, with no significant difference between groups. For example, for pregnant women in the Mamma Mia trial, an automated Web-based PPD prevention intervention, the average number of sessions was 7.4 (of 16) [19]. The e-MB trial, another Web-based individual intervention for PPD prevention, found reasonable adherence (40.47 min of intervention use) but did not have a significant impact in the development of PPD [20]. In the MMB program, an online individual intervention utilizing a mindfulness-based cognitive therapy approach, 57% of women completed at least 4 out of 8 sessions, with the average number of sessions completed at 4.72 [21]. As discussed in the development of the NetMums online PPD treatment intervention, perinatal women have specific needs that are often overlooked in Web-based CBT programs focusing on MDD [44]. Content in Sunnyside was carefully designed to provide desired pregnancy information as well as mood management material, and it is possible that this helped draw women back to the site.

Usability scores indicate that the overall user experience was satisfactory. However, limited uptake of the peer network features suggest that changes could be made to increase group discussion and cohesion. Increased utilization of peer support features might lead to greater differences between groups but given that both groups showed adequate adherence and positive outcomes, optimization of those features may not provide any further benefit.

Although dropout from assessments was high from baseline to 6 weeks postpartum, we retained 72% of participants in assessments through the completion of the intervention. Retention at follow-up dropped considerably, in part because of the time demands on the new mothers. In future trials, this problem will have to be addressed by requiring less time for assessment and doing a better job of preparing participants for the postpartum assessments.

Strengths and Limitations

There are several strengths in this trial. To our knowledge, this is the first online intervention for PPD prevention that included peer social support. Social support is consistently found to be beneficial for pregnant women but can be difficult to access for many women. Women consistently utilized the intervention, even without human support, suggesting that this modality of intervention is appealing to pregnant women.

The limitations of this trial include small sample size, limited diversity in the sample, and the short duration of follow-up. Ideally, follow-up would be closer to 6 months or 1 year to determine if PPD developed later in the postpartum period. Results should be interpreted with caution because of the small sample size, and a larger trial with a more diverse group is needed to verify the outcomes. In addition, most of our participants were recruited via electronic methods, such as email or Research Match, the online volunteer registry. This suggests participants were comfortable using technology. The results may not generalize to those without familiarity with the internet.

Conclusions

In conclusion, this study outlines the development process and feasibility testing of a Web-based intervention to prevent PPD. Results suggest that women were responsive to the intervention although it would benefit from continued refinement. Next steps include a larger trial with a more diverse sample and a longer follow-up period. In addition, making intervention enrollment available via primary care clinics or ob-gyn offices, rather than simply online, may increase access for those most in need of preventive services but with less familiarity with online resources.

Acknowledgments

The study team would like to thank David Mohr, PhD, and Kristina Pecora, PsyD, for their assistance in this research. This study was supported by grants from the NIMH: 1R34MH102478 (JD) and 1R34MH102469 (MWO). Additional support was provided by the National Center For Advancing Translational Sciences of the National Institutes of Health (NIH) under Award Number U54TR001356. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author Note: RG was affiliated with the Department of Psychological and Brain Sciences at the University of Iowa at the time of the work, and is currently affiliated with the Serious Mental Illness Treatment Resource and Evaluation Center and the University of Michigan, Department of Psychiatry. HH was affiliated with the Department of Psychological and Brain Sciences at the University of Iowa at the time of the work, and is currently a graduate student at Western Carolina University. NG was affiliated with the Department of Psychological and Brain Sciences at the University of Iowa at the time of the work, and is currently a graduate student at Kansas State University.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Intervention description.

[PDF File (Adobe PDF File), 85KB - mental_v6i5e10778_app1.pdf]

Multimedia Appendix 2

CONSORT-eHealth Checklist (V 1.6.1).

[PDF File (Adobe PDF File), 513KB - mental_v6i5e10778_app2.pdf]

References

1. O'Hara MW, Wisner KL. Perinatal mental illness: definition, description and aetiology. Best Pract Res Clin Obstet Gynaecol 2014;28(1):3-12. [doi: 10.1016/j.bpobgyn.2013.09.002] [Medline: 24140480]
2. O'Hara MW. Postpartum depression: what we know. J Clin Psychol 2009;65(12):1258-1269. [Medline: 19827112]
3. Field T, Sandberg D, Garcia R, Vega-Lahr N, Goldstein S, Guy L. Pregnancy problems, postpartum depression and early mother-infant interactions. Dev Psychol 1985;21:1152-1156. [doi: 10.1037/0012-1649.21.6.1152]
4. Goodman S, Brand S. In: Hersen H, Gross AM, editors. Handbook of clinical psychology, Vol 2: Children and adolescents. Hoboken, NJ: John Wiley & Sons; 2008:937-965.
5. Hay DF, Pawlby S, Angold A, Harold GT, Sharp D. Pathways to violence in the children of mothers who were depressed postpartum. Dev Psychol 2003;39(6):1083-1094. [Medline: 14584986]
6. Sockol LE, Epperson CN, Barber JO. A meta-analysis of treatments for perinatal depression. Clin Psychol Rev 2011;31(5):839-849. [Medline: 21545782]
7. Sockol LE, Epperson CN, Barber JP. Preventing postpartum depression: a meta-analytic review. Clin Psychol Rev 2013 Dec;33(8):1205-1217 [FREE Full text] [doi: 10.1016/j.cpr.2013.10.004] [Medline: 24211712]
8. Sockol LE. A systematic review of the efficacy of cognitive behavioral therapy for treating and preventing perinatal depression. J Affect Disord 2015 May 15;177:7-21. [doi: 10.1016/j.jad.2015.01.052] [Medline: 25743368]
9. Byatt N, Levin LL, Ziedonis D, Moore Simas TA, Allison J. Enhancing participation in depression care in outpatient perinatal care Settings: a systematic review. Obstet Gynecol 2015 Nov;126(5):1048-1058 [FREE Full text] [doi: 10.1097/AOG.0000000000001067] [Medline: 26444130]
10. Segre LS, Stasik SM, O'Hara MW, Arndt S. Listening visits: an evaluation of the effectiveness and acceptability of a home-based depression treatment. Psychother Res 2010 Nov;20(6):712-721 [FREE Full text] [doi: 10.1080/10503307.2010.518636] [Medline: 21154029]
11. Dennis CL, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536X.2006.00130.x] [Medline: 17150072]
12. Levy L, O'Hara MW. Psychotherapeutic interventions for depressed, low-income women: a review of the literature. Clin Psychol Rev 2010 Dec;30(8):934-950. [doi: 10.1016/j.cpr.2010.06.006] [Medline: 20678834]
13. Mohr DC, Ho J, Duffecy J, Baron KG, Lehman KA, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010 Apr;66(4):394-409 [FREE Full text] [doi: 10.1002/jclp.20659] [Medline: 20127795]
14. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
15. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
16. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
17. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. BMC Psychiatry 2013;13:113. [doi: 10.1186/1471-244X-13-113]
18. Rozental A, Magnusson K, Boettcher J, Andersson G, Carlbring C. For better or worse: An individual patient data meta-analysis of deterioration among participants receiving Internet-based cognitive behavior therapy. J Consult Clin Psychol 2017 Feb;85(2):160-177. [doi: 10.1037/ccp0000158] [Medline: 27775414]
19. Haga SM, Drozd F, Brendryen H, Slinning K. Mamma mia: a feasibility study of a web-based intervention to reduce the risk of postpartum depression and enhance subjective well-being. JMIR Res Protoc 2013 Aug 12;2(2):e29 [FREE Full text] [doi: 10.2196/resprot.2659] [Medline: 23939459]
20. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: a pilot randomized controlled trial. Internet Interv 2015 Sep 01;2(3):257-265 [FREE Full text] [doi: 10.1016/j.invent.2015.06.002] [Medline: 26273567]
21. Dimidjian S, Goodman SH, Felder JN, Gallop R, Brown AP, Beck A. An open trial of mindfulness-based cognitive therapy for the prevention of perinatal depressive relapse/recurrence. Arch Womens Ment Health 2015 Feb;18(1):85-94. [doi: 10.1007/s00737-014-0468-x] [Medline: 25298253]
22. Duggan M, Lenhart A, Lampe C, Ellison NB. Pew Research Center. 2015. Parents and Social Media URL: http://www.pewinternet.org/2015/07/16/parents-and-social-media/ [accessed 2019-02-22] [WebCite Cache ID 76OAxbIrc]

23. Lattie E, Ho J, Sargent E, Tomasino KN, Smith JD, Brown CH, et al. Teens engaged in collaborative health: the feasibility and acceptability of an online skill-building intervention for adolescents at risk for depression. Internet Interv 2017 Jun;8:15-26 [FREE Full text] [doi: 10.1016/j.invent.2017.02.003] [Medline: 28584734]
24. Lara MA, Tiburcio M, Aguilar Abrego A, Sánchez-Solís A. A four-year experience with a web-based self-help intervention for depressive symptoms in Mexico. Rev Panam Salud Publica 2014;35(5-6):399-406 [FREE Full text] [Medline: 25211568]
25. Rollman BL, Herbeck Belnap B, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. JAMA Psychiatry 2018 Jan 01;75(1):56-64 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
26. Houston TK, Cooper LA, Ford DE. Internet support groups for depression: a 1-year prospective cohort study. Am J Psychiatry 2002 Dec;159(12):2062-2068. [doi: 10.1176/appi.ajp.159.12.2062] [Medline: 12450957]
27. Duffecy J, Sanford S, Wagner L, Begale M, Nawacki E, Mohr DC. Project onward: an innovative e-health intervention for cancer survivors. Psychooncology 2013 Apr;22(4):947-951 [FREE Full text] [doi: 10.1002/pon.3075] [Medline: 22438297]
28. Tomasino KN, Lattie EG, Ho J, Palac H, Kaiser SM, Mohr DC. Harnessing peer support in an online intervention for older adults with depression. Am J Geriatr Psychiatry 2017 Oct;25(10):1109-1119 [FREE Full text] [doi: 10.1016/j.jagp.2017.04.015] [Medline: 28571785]
29. Creswell J. Qualitative inquiry & research design: Choosing among five approaches. Thousand Oaks, CA: Sage; 2012.
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
31. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
32. Beck JS. Cognitive therapy: Basics and beyond. New York, New York: Guilford Press; 1995.
33. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. J Med Internet Res 2010 Dec 19;12(5):e48 [FREE Full text] [doi: 10.2196/jmir.1370] [Medline: 21169164]
34. Lund AM. Measuring usability with the USE Questionnaire. STC Usability Newsletter 2001;8(2):3-6 [FREE Full text]
35. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [Medline: 14399272]
36. Potts MK, Daniels M, Burnam MA, Wells KB. A structured interview version of the Hamilton Depression Rating Scale: evidence of reliability and versatility of administration. J Psychiatr Res 1990;24(4):335-350. [Medline: 2090831]
37. Watson D, O'Hara MW, Simms LJ, Kotov R, Chmielewski M, McDade-Montez EA, et al. Development and validation of the Inventory of Depression and Anxiety Symptoms (IDAS). Psychol Assess 2007 Sep;19(3):253-268. [doi: 10.1037/1040-3590.19.3.253] [Medline: 17845118]
38. Kroenke K, Spitzer R. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
39. Kroenke R, Spitzer L, Williams BW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
40. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur Psychiatry 1997;12(5):224-231 [FREE Full text] [doi: 10.1016/S0924-9338(97)83296-8]
41. Tandon D, Perry D, Mendelson T, Kemp K, Leis J. Preventing perinatal depression in low-income home visiting clients: a randomized controlled trial. J Consult Clin Psychol 2011 Oct;79(5):707-712. [doi: 10.1037/a0024895] [Medline: 21806298]
42. Le H, Perry D, Stuart E. Randomized controlled trial of a preventive intervention for perinatal depression in high-risk Latinas. J Consult Clin Psychol 2011 Apr;79(2):135-141. [doi: 10.1037/a0022492] [Medline: 21319897]
43. Verkerk GJ, Pop VJ, Van Son MJ, Van Heck GL. Prediction of depression in the postpartum period: a longitudinal follow-up study in high-risk and low-risk women. J Affect Disord 2003 Nov;77(2):159-166. [Medline: 14607393]
44. O'Mahen H, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]

Abbreviations

CBT: cognitive behavioral therapy
DSM: Diagnostic and Statistical Manual of Mental Disorders
HDRS: Hamilton Depression Rating Scale
ICTS: Institute for Clinical and Translational Science
IDAS: Inventory of Depression and Anxiety Symptoms
ISGs: internet support groups
MDD: major depressive disorder

MINI: Mini International Neuropsychiatric Interview
NIH: National Institutes of Health
PHQ-8: Patient Health Questionnaire-8
PHQ-9: Patient Health Questionnaire-9
PPD: postpartum depression
RCT: randomized controlled trial
SCID-I: Structured Clinical Interview for DSM Axis-I Disorders
UI: University of Iowa

Edited by J Torous; submitted 30.04.18; peer-reviewed by D Tandon, J Felder, A Silva Almodovar; comments to author 21.06.18; revised version received 04.09.18; accepted 30.12.18; published 28.05.19.

Please cite as:
Duffecy J, Grekin R, Hinkel H, Gallivan N, Nelson G, O'Hara MW
A Group-Based Online Intervention to Prevent Postpartum Depression (Sunnyside): Feasibility Randomized Controlled Trial
JMIR Ment Health 2019;6(5):e10778
URL: http://mental.jmir.org/2019/5/e10778/
doi:10.2196/10778
PMID:31140443

Viewpoint

When All Else Fails, Listen to the Patient: A Viewpoint on the Use of Ecological Momentary Assessment in Clinical Trials

Aaron M Mofsen[1], DO; Thomas L Rodebaugh[2], PhD; Ginger E Nicol[1], MD; Colin A Depp[3], PhD; J Philip Miller[4], AB; Eric J Lenze[1], MD

[1]Department of Psychiatry, School of Medicine, Washington University in St Louis, St Louis, MO, United States
[2]Department of Psychological and Brain Sciences, Washington University in St Louis, St Louis, MO, United States
[3]Department of Psychiatry, University of California - San Diego, San Diego, CA, United States
[4]Division of Biostatistics, School of Medicine, Washington University in St Louis, St Louis, MO, United States

Corresponding Author:
Colin A Depp, PhD
Department of Psychiatry
University of California - San Diego
9500 Gilman Drive
San Diego, CA, 92093
United States
Phone: 1 8588224251
Email: cdepp@ucsd.edu

Abstract

A major problem in mental health clinical trials, such as depression, is low assay sensitivity in primary outcome measures. This has contributed to clinical trial failures, resulting in the exodus of the pharmaceutical industry from the Central Nervous System space. This reduced assay sensitivity in psychiatry outcome measures stems from inappropriately broad measures, recall bias, and poor interrater reliability. Limitations in the ability of traditional measures to differentiate between the trait versus state-like nature of individual depressive symptoms also contributes to measurement error in clinical trials. In this viewpoint, we argue that ecological momentary assessment (EMA)—frequent, real time, in-the-moment assessments of outcomes, delivered via smartphone—can both overcome these psychometric challenges and reduce clinical trial failures by increasing assay sensitivity and minimizing recall and rater bias. Used in this manner, EMA has the potential to further our understanding of treatment response by allowing for the assessment of dynamic interactions between treatment and distinct symptom response.

(JMIR Ment Health 2019;6(5):e11845) doi:10.2196/11845

KEYWORDS

ecological momentary assessment; mental health; controlled clinical trial; psychiatry; health technology

Introduction

Background

Mental health treatment development and testing has been at an impasse for the past several decades; our clinical trials increasingly fail more often than in other fields [1]. Although the global burden of psychiatric illness continues to be one of the largest contributors to disability worldwide, investment in the discovery of novel pharmacologic agents flows instead toward disease states with identifiable biological targets. These targets remain elusive in psychiatric disorders [2,3]. The central nervous system (CNS) drug development pipeline has become increasingly burdened with late-phase failures [4], contributing to a well-publicized exodus of the pharmaceutical industry from the CNS space. This has resulted in decreased investment in drug discovery [5].

Treatment Failures: Bad Medicine or Bad Measures?

The randomized, placebo-controlled trial is still considered the gold standard test of treatment efficacy. However, over the past 60 years of treatment research in psychiatry, we have observed that treatment effect sizes remain stable, whereas placebo responses rise [6]. Modern clinical trials are difficult to conduct and are fraught with numerous challenges related to cost, regulatory requirements, recruitment difficulties, and other inefficiencies [7,8]. Added to these challenges is the use of imprecise outcome measures, which hinders the ability to detect true separation of active treatment from placebo response [9].

The contribution of poor measures to treatment failures is particularly well-illustrated in antidepressant trials [10-12]. For example, lanicemine, an N-methyl-D-aspartate receptor antagonist differing from ketamine that produces lower psychotomimetic side effects, was thought to show promise in treating depression [13]. Early phase clinical trials showed promising results in rapidly reversing symptoms of treatment resistant depression, but investigators failed to replicate the results in a late phase study [14]. Similarly, basimglurant, a postsynaptic metabotropic glutamate subtype 5 receptor antagonist, showed promise in early phase trials but failed to separate from placebo on the primary outcome measure in a larger phase 2b trial [15]. In both cases, the primary end point was change from baseline to 6 weeks in the Montgomery Asberg Depression Scale (MADRS), which is considered an industry standard in depression treatment research. The authors identified flaws in study design, conduct, and even underlying scientific rationale as possible causes of these late stage failures.

It seems unlikely, given the financial and intellectual resources brought to bear in the early phases of discovery, that investigators could have gotten the scientific rationale so wrong. A more probable explanation for the failed studies might lie in how the primary outcome was determined and measured. Although the MADRS is considered a standard assessment tool in depression research, poor interrater reliability (ie, imprecision of measurement) is one of many limitations to this measure's assay sensitivity.

The Culprit: Faulty Signal Detection

Measurement *assay sensitivity*, as it applies to clinical research, refers to the ability of a symptom assessment measure to detect whether a difference exists between treatment groups [16]. Issues of assay sensitivity are well known in psychiatric treatment research and have been observed with older self-report scales such as the Hamilton Rating Scale for Depression (HAM-D) as well as in newer clinician-administered instruments such as the MADRS. Both measures include several symptom domains but offer only a final summed score. This offers little insight into the specific symptoms underlying the clinical presentation.

Self-report measures may incorporate reporter bias, whereas clinician-administered assessments incorporate bias on the part of the clinician. For example, there may be bias in recruitment or sample ascertainment, such as career patients who serially enroll in research studies for financial reasons and are thus motivated to answer questions in such a way as to increase likelihood of enrollment. Investigators may unconsciously inflate baseline measures of psychiatric symptoms to meet recruitment goals [17-19].

Nonetheless, these arguments fail to explain why academic studies, in which less financial gain accrues to the patient and investigator, also see a high placebo response and failure rate [20]. Regardless, reduced assay sensitivity in clinical trials has the potential to sabotage treatment development at any stage. We submit that these and other depression symptom measures reduce assay sensitivity in 3 primary ways: unnecessary complexity, human error (ie, clinician judgment), and infrequent sampling.

Getting to Precision Assessment

The idea of using technology to increase the accuracy and precision of symptom assessment in clinical trials is gaining momentum. For example, the National Institutes of Health toolbox was designed specifically for this purpose [21]. The Patient-Reported Outcomes Measurement Information System also offers researchers standardized patient-reported outcome (PRO) measurement tools with transparent performance metrics [22]. Self-report measures delivered via mobile technology certainly offer ecological validity and may also prove superior to clinician-administered instruments in large, industry-funded clinical trials. Improved measurement would likely translate into more useful clinical trials. It may even go a long way toward surmounting our present impasse in developing new mental health treatments.

Clearly, we are not the first to contemplate the problem of assay sensitivity in our field. However, public discussion as to why progress in the field of psychometrics has stalled has not extended to industry trials. Open scientific discourse has also been limited on the subject of developing novel, effective, Food and Drug Administration (FDA)–sanctioned instruments, which could be used to track mental health disorder outcomes with greater assay sensitivity. As the success or failure of antidepressant treatment trials often rests solely on the presumed validity and reliability of symptom measures, it should follow that these assessments deserve the same degree of scrutiny regarding assay sensitivity as any laboratory test.

In this viewpoint, we will examine 3 major problem areas we believe the field needs to address in getting to precision assessment: overly complex assessment tools, contributions of human error, and limitations of infrequent sampling. First, we will review the 2 gold standard depression instruments used at present to track psychiatric symptoms in industry-funded drug trials. Next, we will examine the role of clinician assessment and how human involvement in measurement contributes to error. We will then discuss challenges to adequate measurement frequency in obtaining valid self-report data. Finally, we propose a solution to the measurement problem in depression clinical trials. We will explore contributions from the fields of mathematics, human psychology, and computer science to the development of mobile technology–based measures, which we believe may offer significant improvements over traditional symptom assessment.

Problem 1: Needless Complexity Undermines Utility

Key point:

- Overly broad measures that attempt to cover multiple symptoms or symptom domains compromise signal detection. To meaningfully reduce error, consensus on what to measure is needed.

The Problem of Excessive Description

Psychiatric rating scales frequently use diagnostic criteria or descriptive psychopathology to track a patient's progress throughout a clinical trial. The descriptive psychopathology for a given psychiatric disorder is by nature more expansive than the diagnostic criteria alone, which can be helpful for identifying clinically significant features for treatment targets. This problem is not restricted to mental health research; trials in cardiology have also been compromised by failing to adequately confine outcome measures for meaningful signal detection [23]. In major depression, patients often have irritability, anxiety, and other symptoms in addition to the 9 cardinal symptoms of the disorder. A content analysis by Eiko Fried found 52 symptoms of depression across 7 commonly used depression scales, with a content overlap among all scales of only 32 percent [24].

Take for example the MADRS discussed above [25]. The clinician in using this scale administers a 10-item assessment to a study participant. The change in the total score over time is then used to determine whether the treatment under investigation is effective. The 17-item HAM-D (HAM-D-17) determines efficacy similarly [26]. However, both items assess multiple symptom domains, all considered diagnostic aspects of depression. A recent study by Checkroud et al [27] of over 7000 patients with major depression demonstrates why this approach, as well as any other that relies on indiscriminate use all of the items in a scale to assess primary efficacy outcomes (eg, the HAM-D), may be a problem. In their study, they illustrate how this indiscriminate approach to measurement can jeopardize a potential treatment in late-phase clinical trials. Specifically, they found that consistent antidepressant treatment response was found only for the core emotional symptoms (anergia, dysphoria, anhedonia, feelings of worthlessness, and difficulty concentrating). The detectable signal for treatments shown to be effective is thus obscured by the total score, which is the only score considered when designing trials to determine efficacy. This example highlights how standard rating scales have contributed to treatment failures by introducing unnecessary *complexity*, which reduces measurement specificity.

To further complicate matters, measuring multiple constructs inflates the chance that items tied to each construct will shift unpredictably over time (eg, due to lack of longitudinal factorial invariance) [28]. In this way, depression rating scales are often a mix of sensitive and specific items (dysphoria, anhedonia), nonspecific items (anxiety), and symptoms that may be derived from an unrelated illness (eg, fatigue). Side effects of the treatment itself are also frequently conflated with the items in the primary outcome measure. Moreover, individual items within a scale are often not weighted for relevance. As the success or failure of a treatment rests on a scale's summative score, it follows that some of the score's equally weighed items might be totally irrelevant to the trajectory of the disorder in question [29]. The 24-item HAM-D (HAM-D-24) is more comprehensive than the 17-item version [30]. It was designed to more comprehensively capture relevant symptoms. However, using the HAM-D-24 may conceal treatment effects by introducing items that assess uncommon or diagnostically nonspecific symptoms, such as hypochondriasis or depersonalization. Again, as the total score is used to determine whether or not a treatment is effective, there is a further risk of magnifying irrelevant changes and obscuring important ones.

Less is More

The shortened 6-item HAM-D and MADRS scales, which favor core items such as low mood, anhedonia, and guilt, have both been shown to be more sensitive than HAM-D-17 and the 10-item MADRS, respectively [31]. The shorter 6-item version of the HAM-D [32] was superior to the longer HAM-D-17, 21 and 24 in detecting treatment response to the newer antidepressant vortioxetine versus placebo [33]. Similarly, the buprenorphine/samidorphan combination treatment, which failed to separate from placebo on the primary outcome measure of change from baseline on the MADRS-10 item scale, fared better in separating from placebo using the MADRS-6 item scale [34]. These examples suggest a data reduction approach to symptom assessment focusing on core symptoms is more likely to accurately detect meaningful clinical response. Unfortunately, there is, as of yet, little agreement on which symptoms are most relevant.

Consensus on the most clinically, functionally, or personally relevant features of treatment response or remission is needed to improve signal detection. If we simply wish to use our existing scales more pragmatically, we would take a treatment we know to be effective and choose the individual items from a selected scale that reveal the greatest amount of separation in favor of the proven treatment. We would then use the items from that same scale to determine whether or not an unproven treatment is effective. Alternatively, the field could adopt a universal consensus around measuring the core emotional symptoms of the illness to determine treatment success or failure. This is a difficult and unlikely scenario as we do not have the evidence base at present necessary to establish what exactly these core symptoms might be. In either case, improvement from a functional or pharmacoeconomic perspective may not map well onto any of the items in the measures we currently use. This may force the field to revisit some of its a priori assumptions about clinical relevance. In short, although we can confidently say that our current approach is suboptimal, fixing it will not be so easy.

Problem 2: Human Error Magnifies Measurement Error

Key points:

- Clinician-administered scales compound response bias
- Self-report alone is imperfect but minimizes rater contribution to measurement error

Not All That Glitters is Gold

Psychiatric treatment research has traditionally considered clinician-administered assessments to be the *gold standard* over PRO measures. This stems in part from an inherent belief that the clinician *objectively corrects* for whatever error (eg, errors of omission, exaggeration, expectancy effect, and Hawthorne effect), intentional or otherwise, introduced by the patient. Perhaps somewhat counterintuitively, clinicians may *magnify* the patient's error. A large study evaluating self-report and clinician-administered instruments from the Sequenced Treatment Alternatives to Relieve Depression trial found that self-report measures contributed more to the prediction of outcomes of clinician-administered instruments than vice versa [35]. The authors of the study also recommended that, in the event that only 1 form of assessment could be used, self-reported outcome measures would be preferable.

Error or bias on the part of the clinician is routine, rather than idiosyncratic. It would be unfair to presume it to be the result of malice or laziness. It may happen unconsciously and even in good faith because clinical judgment is not completely objective. Interviewers are also susceptible to either a positive or negative rater bias depending on whether research participant attributes, often irrelevant to the assessment at hand, are perceived as positive or negative. This can result in sometimes pronounced unconscious alterations of judgment [36] that significantly impact clinical decision making. This has been illustrated in studies finding poor interrater and test-retest reliability in standard clinician-administered assessment measures for depression [3]. The reason for such results may be that clinicians, even when given rules governing the scoring of the assessment at hand, will tend to drift from standard calibrated practice [37]. Whether or not a clinician reliably follows an assessment-related rule depends on the amount of inertia that must be overcome to adopt it, the format in which the rule was originally presented, the number of demands that compete with the rule, and the institutional pressures involved in maintaining compliance with the rule [38].

When all Else Fails, Listen to the Patient

Although the evidence is still far from conclusive, a decent body of literature has elevated the stature of PROs vis-a-vis traditional, clinician-administered rating scales. Self-report assessments represent an improvement over clinician-administered assessments insofar as they eliminate rater bias and reduce the likelihood that participants will feel compelled to give socially desirable responses (a type of response bias) or affirmative answers when interviewed face-to-face [39]. For example, a large meta-analysis of placebo response in 96 antidepressant trials by Mora et al found that clinician-administered instruments were associated with a higher placebo response than PRO measures [40]. Such evidence further supports the idea that clinician-administered scales add error rather than removing or mitigating patient error. In summary, although we place a high value on clinician-administered assessments, clinician objectivity may be more of an appealing myth than reality.

Problem 3: Infrequent Sampling Hurts Sensitivity

Key points:

- Retrospective patient symptom report in the context of a clinical trial may be inaccurate
- Ecologically valid symptom reports collected in real time are needed to interpret treatment effects

(Not So) Total Recall

Self-report also has inherent limitations. This was recognized by Arthur Schopenhauer in the 19th century [41], who observed that one cannot be both the subject and object of accurate perception. Thus, reporting on one’s own mood even in the present poses significant challenges and represents an irremediable layer of error. Mehl and Conner have also comprehensively discussed the problem of recall bias in psychological research [42]. In short, asking a participant to provide a retrospective symptom report merely compounds this error by introducing recall bias. In other words, emotional recall bias (unlike the subject-object problem) is a controllable source of error. Neuroscientists have found memory to be frequently unreliable, particularly when the encoding and retrieval of memories occurs during periods of emotional arousal [43]. Memory has many odd biases, not all of which are evident in daily life. For instance, it has been shown that people have a tendency to remember events that ought to be enjoyable, such as a vacation or spending time with one’s children, as being more pleasant than they actually were [42]. Thus, asking a respondent to recall something requires filtration through whatever emotional state the subject happens to be in at the time of the assessment, which only compounds this error [44]. Furthermore, respondents are unlikely to accurately create a coherent summary of their emotional states over time.

What is the (Right) Frequency?

Infrequent measurement or sampling in clinical trials tacitly makes the assumption that we know enough about how an illness behaves over time to ask questions with a time frame modifier (eg, “In the last week...”) and is associated with measurement error in clinical trials. This has been illustrated in disciplines outside of psychiatry. For example, the Heart Outcomes Prevention Evaluation trial evaluated the effect of the angiotensin-converting enzyme inhibitor ramipril in patients at high risk for adverse cardiovascular events [45]. The study found that ramipril lowered blood pressure assessed via 24-hour ambulatory measurement, whereas office-based blood pressure measurements did not detect the treatment response. Investigators attributed this to a diurnal variation in blood pressure or *white coat hypertension* —phenomena that could not be captured with the limited number of measures obtained during office hours or that were affected by the office visit itself. For this reason, blood pressure assessment in clinical trials has moved to using frequent ambulatory blood pressure sampling to assess treatment efficacy, which has essentially eliminated the placebo response in antihypertensive treatment trials [46,47].

Similar to blood pressure, depressive symptoms also appear to fluctuate throughout the day or in response to specific situations [48]. Mobile technology offers a feasible way to increase sampling frequency, as evidenced by the already rich scientific literature on ambulatory assessment [42]. However, this approach has yet to be fully embraced by industry sponsored studies, where it could be of prime utility. To date, only 1 industry-sponsored study currently underway has attempted to compare daily, ambulatory self-report with a clinician-administered measure [49]. Frequent, in-the-moment self-report also has its limitations. There is no doubt some theoretical limit on high-frequency sampling to the extent that it may, if administered often enough, conflate mood and emotions or succeed in becoming itself a source of negative mood, affect, or emotions [50,51]. However, this issue calls for careful experimentation with frequency to assess acceptability rather than avoiding frequent sampling altogether.

The State Versus Trait Problem

Symptoms of many psychiatric illnesses are characterized as trait-like in advance of any evidence to support this assumption. However, variation is routinely observed in behaviors studied over time, irrespective of how trait-like they seemed to be (eg, personality traits such as sociability) [52]. For this reason, it is highly probable that important variation is the rule rather than the exception in psychiatric illness. For example, in an individual with major depression, mood might be very depressed at a certain point in the morning and near-normal later that same day [48].

Despite this, we continue to measure mood as a stable trait-like symptom (eg, “in the last 7 days, how has your mood been?”). This is the case for most psychiatric symptom assessments, where dynamic versus stable or trait-like nature of symptoms are poorly described. The only way to ascertain variation or lack thereof is to sample the illness frequently *before finalizing the measure (eg, for use in a treatment study)*. In other words, frequent sampling would ideally be used to inform the creation of a scale before using it to track efficacy [52]. Without this approach, scale selection becomes thoughtlessly reflexive [50]. Limited sampling likely further compromises psychiatric research because trait measures require respondents to attempt a summation of states via recall of past experiences, which has been shown to introduce error [53].

Even if the symptoms of psychiatric illness are predominantly trait-like, we would continue to favor frequent sampling, even if this requires us to use a smaller number of items. This is in contrast to classical test theory, from which we take the maxim that adding equally good items to a measure leads to greater reliability and therefore, a better shot at validity [54]. This is based on the ideal circumstance where it is possible to ask a respondent the same question repeatedly, which we cannot do at a single time point without expecting the respondent to become reactive to the question [54,55]. Furthermore, a measure using high-signal items repeatedly over time would better capture any given quality than would a measure with a mix of items with lower signal detection at a single time point [56]. In psychiatric treatment research, we have historically chosen to use a greater number of inferior items at a single time point, even though the maxim we are following was based on equations that are arguably better suited to repeated measurement of a single quality.

Solution: Ecological Momentary Assessment

Overview

Ecological momentary assessment (EMA) is frequent, real time, patient-reported assessment delivered via surveys (eg, “right now, my mood is...”) and completed by the patient typically via mobile device to collect information about the patient in a real-world setting [57]. Participants are prompted at prespecified intervals to complete symptom assessments rather than having a prompt dependent upon a passive event (eg, actigraphy and patterns of speech). EMA may overcome the deficiencies inherent in traditional clinician-administered instruments. Evidence from pain studies examining EMA alongside retrospective recall show a consistent discrepancy between the 2 forms of report [58]. A similar discrepancy between real time and retrospective self-report of affect has also been demonstrated [59]. A single item scale measuring mood delivered via EMA outperformed the HAM-D-17 in its ability to predict “current relapse status” in patients with major depressive disorder [60].

Increasing Accuracy in Early Phase Trials

Frequent, real-time EMA sampling has been shown in the same study to both qualify positive findings in clinical trials and detect treatment effects that the HAM-D was unable to detect between groups after 18 weeks of treatment [61]. Frequent real-time sampling has also been shown to unmask differences between treatment responders and nonresponders and to detect treatment effects earlier than clinician-administered assessments [62,63]. Finally, frequent, real-time sampling compared with retrospective assessment has been shown to increase the precision of measurement over time.

An example of how infrequent sampling adversely affects assay sensitivity in clinical trials was recently provided by Moore et al [64]. In this study, the researchers assessed the effects of mindfulness-based stress reduction (MBSR), compared with an attention placebo. For outcome assessments, they measured depressive symptoms, anxiety symptoms, and mindfulness self-ratings in 2 ways: EMA tools delivered to participants electronically via a smartphone 3 times daily for 14 days and traditional paper- and pencil-based measurement tools asking about last week’s symptoms (comparable with most outcome measures). The EMA-based outcome assessment resulted in a much lower number needed to treat (NNT) for MBSR than the same outcomes measured using the traditional technique: the NNT for treating depression was 8 using EMA versus 31 using traditional measurement. In other words, EMA captured a treatment effect that was missed by standard self-report assessments. This was also reflected in the smaller SDs for outcomes measured via EMA when averaged over time. In short, frequent ambulatory assessment improves precision.

Increased Understanding of Core Symptom Constructs

EMA may also increase measurement precision by tracking how symptoms of an illness behave and interact over time [65]. This allows investigators to characterize state versus trait-like symptoms and establish the nature of the relationships between symptoms over time. This approach may also be useful because it offers the ability to evaluate interactions between symptoms without first assuming that they are symptoms *of the disorder in question*. This “pragmatic nihilism” [66] or “symptomic” [67] approach differs from how we currently assess psychiatric disorders. Clinician-administered instruments are rated with the built-in assumption that any number of symptoms are all tied to 1 underlying, latent variable (eg, depression). With enough patient-reported EMAs carried out over time, investigators may be able to observe how symptoms interact with one another.

It may also be possible to discern which symptoms are central to the disorder under study and how certain upstream symptoms may influence a cascade of symptoms downstream. How many EMAs are *enough* depends on the exact questions being asked and the assumptions made in the analysis; however, it is likely that as little as 25 measurements from hundreds of participants or a hundred measurements in even a small number of participants would be a reasonable starting place [68]. Such findings may eventually afford researchers the unique opportunity to stratify clinical trial participants based on *how* they do or do not get better rather than simply whether or not they get better. The approach becomes highly descriptive at the level of the individual, thereby allowing one to answer a host of previously unanswerable questions.

Deconstructing Treatment Response

Another question that might be asked is whether patients responding to an intervention or placebo get better in the same way. In other words, do the *temporal dynamics* of placebo response differ from that observed in drug response? Temporal dynamics here refer to certain discernable patterns in the EMA data that allow a researcher to broadly classify a patient as displaying, for instance, affective inertia (symptoms strongly relate to themselves over time, resulting in less change over time), affective instability (symptoms vary a great deal over time), or inability to differentiate between symptoms (as 1 symptom gets better or worse the rest tend to follow) [69]. This is by no means an exhaustive list of questions that may be asked of the data derived from EMA. It is safe to say EMA has the potential to offer a renaissance of sorts in descriptive psychopathology and may even allow for veritable *personalized medicine* given the types of patterns and points of intervention it is able to reveal.

EMA may also help us detect the phenomenon of regression to the mean. This phenomenon occurs when a baseline assessment of symptoms in a clinical research study is inflated at the initial visit before regressing to where those symptoms normally *live*. This is thought to significantly impact the ability to detect separation whenever it occurs in the placebo group. Using EMA, patients may be monitored in the outpatient setting not simply for clinical research purposes but rather to give the clinician a better idea of whether or not a patient is getting better. This approach appreciates EMA as an instrument that may be used to conduct field research, which is thought to have better “ecological validity” than assessments delivered within the artificial environment of the clinical trial site [42]. Such real-world information could be used to find out where that patient “lives” if a patient is being screened for a clinical research study. Similarly, it is not difficult to envision tailoring inclusion/exclusion criteria to this end. If and when this does take place, CNS research will be indebted to data provided directly by the patient.

Developing Better Interventions

Once individual symptom characteristics are known, targeted interventions can be developed. For instance, if insomnia leads to anergia the following day, which in turn leads to anhedonia, one might examine whether applying an intervention at the onset of insomnia changes the observed course of symptomatology downstream. This sort of intervention is called an ecological momentary intervention (EMI) because it relies on EMA or a just-in-time adaptive intervention. An EMI is an intervention informed by data gathered by EMA. We can already find examples of researchers using EMA data to provide an EMI. For example, EMI has already been shown to be very successful in providing patients with substance use disorders relapse prevention tools precisely when they need it the most [70]. It is conceivable that EMA scales, in addition to providing efficacy outcomes with increased assay sensitivity, may also reveal novel points of intervention in clinical trials.

Multiple methods, including multilevel vector autoregression and multilevel dynamic structural equation modeling, can help researchers examine how individuals may vary from group trends over time [71,72]. This might allow clinicians to tailor a personalized EMI based on a patient’s own unique pattern of EMA data. To take this idea further still, EMA may eventually be able to offer the unique ability to evaluate whether a target is being addressed by an intervention via *real-time* lagged mediation rather than post hoc analyses. In other words, we would be able to use real-time lagged mediation to see whether or not we are actually engaging a chosen target precisely when we are attempting to target it.

The use of EMA to gather the data needed to deliver a just-in-time EMI is also consistent with the concept of target engagement raised by the National Institute of Mental Health in an effort to address the declining success of clinical trials in mental health. A target is defined as something “molecular, cellular, circuit, behavioral or interpersonal, commensurate with the intervention,” which is expected to be changed in some way by the intervention being studied [73]. The concept of target engagement is closely related to a recent call for a research focus on symptomics or the examination of “symptom-specific effects” [70]. Such a focus, as represented in the example above, may allow us to identify those key symptoms that tend to precede or perhaps even cause other symptoms. Investigating patterns of interaction between symptoms in this way may help us to understand some of the underlying causes of complex psychiatric illnesses.

How Do We Get to Widespread Use of Ecological Momentary Assessment in Clinical Trials?

Understanding and Getting Past Limitations

Although smartphone ownership is not universal, it is increasing, particularly among individuals with psychiatric conditions. John Torous found in a recent survey of 457 individuals with schizophrenia or schizoaffective disorder that greater than half (54%) of such individuals owned a smartphone [74]. Perhaps a greater question then is whether a participant with a smartphone would want to use it to regularly quantify his or her depressive symptoms. User privacy is also becoming an increasingly important issue as faith in *big tech* to safeguard users’ privacy has waned in the wake of the numerous scandals. Getting around these limitations may require sponsors to invest in low-cost devices participants can use while enrolled in trials.

Use of EMA in the real world often leads to missing data that have historically made analysis problematic. Users may not be compliant with the number of surveys they are required to complete in a timely manner, and, as described above, frequency of assessments increase precision only up to a point. Beyond this point, with too frequent assessment, the risk increases of either introducing noise by sampling irrelevant aspects of the human condition or of the assessment itself becoming a negative part of the intervention. Investigators will have to consider an assay sensitivity assessment as part of the startup process to determine how the target population will best respond to EMA.

Although the FDA has made its expectations for PRO measures clear [75], it is not at all clear whether every aspect of FDA guidance will neatly translate to electronic PROs. For example, to what extent, if any, would necessary software updates for an accepted EMA app involve the FDA? FDA guidance for evaluating antidepressant drugs has not been updated since 1977 and explicitly favors selecting scales that have been previously used in drug trials over ones that are novel [76]. This effectively prioritizes tradition over innovation and creates a catch-22 for researchers who might otherwise break with the status quo. Clinician-administered instruments need to be evaluated alongside commensurate EMA-delivered items. This will help us to determine parameters such as the optimal sampling frequency but will likely also be necessary as the FDA typically reports correlation coefficients for established measurement tools [77].

The conceptualization of disorders based on Diagnostic and Statistical Manual of Mental Disorders/International Classification of Diseases criteria has been called into question and may eventually be replaced altogether by Research Domain Criteria [78]. Although EMA is in many ways conducive to a dimensional approach to mental illness, this migration would obviously require a new approach to EMA scale creation and validation. In this case, the role of EMA may be to supplement observable behaviors with self-report.

EMA may not be ideal for detecting rare events, especially if they occur infrequently relative to the sampling frequency (ie, as the sampling frequency decreases so too does the probability of capturing *rare events*). Thus, when and how to apply EMA in clinical trials remains an area requiring additional study and consensus development.

EMA should not be mistaken for a panacea so long as p-hacking, publication bias, and alpha inflation continue to affect the integrity of clinical research. Any scale used to evaluate the efficacy of an intervention in large industry-sponsored clinical trials must be uniform and well-validated. Thus, to create a standard efficacy measure for a given psychiatric disorder, we first must form a consensus about the types of items that should be included in the EMA scales, the frequency and duration of assessments, and the types of analytical approaches that will be used to interpret the data. The FDA would be unlikely to accept an EMA-based primary outcome measure over existing efficacy end point measures without standardization across multiple field trials in different populations. These data should then clearly establish test-retest reliability, external validity, and other parameters necessary to validate an EMA scale.

Conclusions

Moving from clinician-administered rating scales toward real-time patient-reported measures such as EMA offers significant advantages across medical settings. In clinical research studies, EMA may reduce placebo response and increase intervention-placebo separation. EMA also offers an obvious advantage over clinician-administered rating scales in inpatient and community settings given that time, cost, and staff pressures make use of the latter measure impractical. In community and inpatient settings, EMA can be used to identify individual factors leading to relapse, provide a more accurate picture of how a patient has been doing between clinical visits, and link real-world functional outcome measures over time (eg, rates of rehospitalization, days lost because of disability, and likelihood of self-harm) to *scores* on EMA scales. Finally, interventions are rapidly being introduced and delivered via smartphone. EMA may offer the best way to assess intervention acceptability and efficacy, creating the opportunity to personalize treatments with real-time adaptation. For these reasons, EMA is poised not only to replace clinician-administered rating scales in research settings but also to increase accessibility of EMA measures to the patients and health care providers in clinical settings, ultimately allowing real-world clinical settings to contribute meaningful data to research and development of new interventions.

Overall, we believe that the continued use of clinician-administered retrospective self-report assessments in clinical trials contributes significantly to observed treatment failures and squanders innovative potential. As we have described, the instruments currently being used are too broad to adequately assess outcomes, suffer from poor interrater reliability, make inappropriate assumptions about how the illness being studied behaves, and rely on patient recall despite a sizeable body of research, which cautions against this. EMA instruments may play an increasingly important role in addressing the disparity between the need for and investment in novel mental health treatments. Self-report assessment via EMA addresses the limitations of traditional assessment methods but has not yet made its way into large multisite clinical trials sponsored by the industry. Although the FDA's recent efforts to advance mobile technology in clinical trials [79] represents an important first step, iterative testing of standardized EMA-delivered instruments to assess primary outcomes in clinical research is still needed.

Conflicts of Interest

None declared.

References

1. Pankevich DE, Altevogt BM, Dunlop J, Gage FH, Hyman SE. Improving and accelerating drug development for nervous system disorders. Neuron 2014 Nov 5;84(3):546-553 [FREE Full text] [doi: 10.1016/j.neuron.2014.10.007] [Medline: 25442933]
2. Hyman SE. Psychiatric drug development: diagnosing a crisis. Cerebrum 2013 Mar;2013:5 [FREE Full text] [Medline: 23720708]
3. Insel TR, Voon V, Nye JS, Brown VJ, Altevogt BM, Bullmore ET, et al. Innovative solutions to novel drug development in mental health. Neurosci Biobehav Rev 2013 Dec;37(10 Pt 1):2438-2444 [FREE Full text] [doi: 10.1016/j.neubiorev.2013.03.022] [Medline: 23563062]
4. Marder SR, Laughren T, Romano SJ. Why are innovative drugs failing in phase III? Am J Psychiatry 2017 Sep 1;174(9):829-831. [doi: 10.1176/appi.ajp.2017.17040426] [Medline: 28859511]
5. Miller G. Is pharma running out of brainy ideas? Science 2010 Jul 30;329(5991):502-504. [doi: 10.1126/science.329.5991.502] [Medline: 20671165]
6. Leucht S, Leucht C, Huhn M, Chaimani A, Mavridis D, Helfer B, et al. Sixty years of placebo-controlled antipsychotic drug trials in acute schizophrenia: systematic review, Bayesian meta-analysis, and meta-regression of efficacy predictors. Am J Psychiatry 2017 Dec 1;174(10):927-942. [doi: 10.1176/appi.ajp.2017.16121358] [Medline: 28541090]
7. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. Lancet 2014 Jan 11;383(9912):176-185 [FREE Full text] [doi: 10.1016/S0140-6736(13)62297-7] [Medline: 24411646]
8. Nutt D, Goodwin G. ECNP Summit on the future of CNS drug research in Europe 2011: report prepared for ECNP by David Nutt and Guy Goodwin. Eur Neuropsychopharmacol 2011 Jul;21(7):495-499. [doi: 10.1016/j.euroneuro.2011.05.004] [Medline: 21684455]
9. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015 Oct;14(3):294-300 [FREE Full text] [doi: 10.1002/wps.20241] [Medline: 26407778]
10. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. J Am Med Assoc 2002 Apr 10;287(14):1840-1847. [doi: 10.1001/jama.287.14.1840] [Medline: 11939870]
11. Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother Psychosom 2003;72(3):115-127. [doi: 10.1159/000069738] [Medline: 12707478]
12. Iovieno N, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. J Clin Psychiatry 2012 Oct;73(10):1300-1306. [doi: 10.4088/JCP.11r07485] [Medline: 23140647]
13. Sanacora G, Johnson MR, Khan A, Atkinson SD, Riesenberg RR, Schronen JP, et al. Adjunctive lanicemine (AZD6765) in patients with major depressive disorder and history of inadequate response to antidepressants: a randomized, placebo-controlled study. Neuropsychopharmacology 2017 Mar;42(4):844-853 [FREE Full text] [doi: 10.1038/npp.2016.224] [Medline: 27681442]
14. Kobak KA, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? J Clin Psychopharmacol 2007 Feb;27(1):1-5. [doi: 10.1097/JCP.0b013e31802eb4b7] [Medline: 17224705]
15. Quiroz JA, Tamburri P, Deptula D, Banken L, Beyer U, Rabbia M, et al. Efficacy and safety of basimglurant as adjunctive therapy for major depression: a randomized clinical trial. JAMA Psychiatry 2016 Jul 1;73(7):675-684. [doi: 10.1001/jamapsychiatry.2016.0838] [Medline: 27304433]
16. Snapinn SM. Noninferiority trials. Curr Control Trials Cardiovasc Med 2000;1(1):19-21 [FREE Full text] [doi: 10.1186/cvm-1-1-019] [Medline: 11714400]
17. Puttagunta PS, Caulfield TA, Griener G. Conflict of interest in clinical research: direct payment to the investigators for finding human subjects and health information. Health Law Rev 2002;10(2):30-32. [Medline: 15739309]
18. McCann DJ, Petry NM, Bresell A, Isacsson E, Wilson E, Alexander RC. Medication Nonadherence, "Professional Subjects," and Apparent Placebo Responders: Overlapping Challenges for Medications Development. J Clin Psychopharmacol 2015 Oct;35(5):566-573 [FREE Full text] [doi: 10.1097/JCP.0000000000000372] [Medline: 26244381]
19. Kobak KA, Leuchter A, DeBrota D, Engelhardt N, Williams JB, Cook IA, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. J Clin Psychopharmacol 2010 Apr;30(2):193-197. [doi: 10.1097/JCP.0b013e3181d20912] [Medline: 20520295]
20. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. Am J Psychiatry 2013 Jul;170(7):723-733 [FREE Full text] [doi: 10.1176/appi.ajp.2012.12040474] [Medline: 23318413]
21. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. Neurology 2013 Mar 12;80(11 Suppl 3):S2-S6 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872e5f] [Medline: 23479538]
22. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. J Appl Meas 2010;11(3):304-314 [FREE Full text] [Medline: 20847477]

23. Pocock SJ, Stone GW. The primary outcome fails - what next? N Engl J Med 2016 Sep 1;375(9):861-870. [doi: 10.1056/NEJMra1510064] [Medline: 27579636]
24. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. J Affect Disord 2017 Dec 15;208:191-197. [doi: 10.1016/j.jad.2016.10.019] [Medline: 27792962]
25. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr;134:382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
26. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
27. Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. JAMA Psychiatry 2017 Apr 1;74(4):370-378 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0025] [Medline: 28241180]
28. Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, Borsboom D. Measuring depression over time...Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. Psychol Assess 2016 Dec;28(11):1354-1367. [doi: 10.1037/pas0000275] [Medline: 26821198]
29. Hieronymus F, Emilsson JF, Nilsson S, Eriksson E. Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. Mol Psychiatry 2016 Apr;21(4):523-530 [FREE Full text] [doi: 10.1038/mp.2015.53] [Medline: 25917369]
30. Guy W. NCDEU Assessment Manual for Psychopharmacology. Washington, DC: US Department of Health, Education, and Welfare; 1976:91-338.
31. Bech P. Rating scales in depression: limitations and pitfalls. Dialogues Clin Neurosci 2006;8(2):207-215 [FREE Full text] [Medline: 16889106]
32. Bech P, Gram LF, Dein E, Jacobsen O, Vitger J, Bolwig TG. Quantitative rating of depressive states. Acta Psychiatr Scand 1975 Mar;51(3):161-170. [Medline: 1136841]
33. Kyle PR, Lemming OM, Timmerby N, Søndergaard S, Andreasson K, Bech P. The validity of the different versions of the Hamilton Depression Scale in separating remission rates of placebo and antidepressants in clinical trials of major depression. J Clin Psychopharmacol 2016 Oct;36(5):453-456. [doi: 10.1097/JCP.0000000000000557] [Medline: 27525966]
34. Carroll J. Endpoint News. 2016. Alkermes plots course to the FDA after its depression drug scores success in last-stand PhIII URL: https://tinyurl.com/y52rk6vm [accessed 2018-08-07] [WebCite Cache ID 71U50UuRV]
35. Uher R, Perlis RH, Placentino A, Dernovšek MZ, Henigsberg N, Mors O, et al. Self-report and clinician-rated measures of depression severity: can one replace the other? Depress Anxiety 2012 Dec;29(12):1043-1049 [FREE Full text] [doi: 10.1002/da.21993] [Medline: 22933451]
36. Nisbett RE, Wilson TD. The halo effect: evidence for unconscious alteration of judgments. J Pers Soc Psychol 1977;35(4):250-256 [FREE Full text] [doi: 10.1037/0022-3514.35.4.250]
37. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. Science 1989 Mar 31;243(4899):1668-1674. [Medline: 2648573]
38. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003 Oct 11;362(9391):1225-1230. [doi: 10.1016/S0140-6736(03)14546-1] [Medline: 14568747]
39. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291 [FREE Full text] [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]
40. Mora MS, Nestoriuc Y, Rief W. Lessons learned from placebo groups in antidepressant trials. Philos Trans R Soc Lond B Biol Sci 2011 Jun 27;366(1572):1879-1888 [FREE Full text] [doi: 10.1098/rstb.2010.0394] [Medline: 21576145]
41. Schopenhauer A. The World As Will And Representation, Volume 2. New York: Dover Publications; 1966.
42. Mehl MR, Conner TS. Handbook of Research Methods for Studying Daily Life. New York: Guilford Press; 2013.
43. Lacy JW, Stark CE. The neuroscience of memory: implications for the courtroom. Nat Rev Neurosci 2013 Dec;14(9):649-658 [FREE Full text] [doi: 10.1038/nrn3563] [Medline: 23942467]
44. Urban EJ, Charles ST, Levine LJ, Almeida DM. Depression history and memory bias for specific daily emotions. PLoS One 2018;13(9):e0203574 [FREE Full text] [doi: 10.1371/journal.pone.0203574] [Medline: 30192853]
45. Svensson P, de Faire U, Sleight P, Yusuf S, Ostergren J. Comparative effects of ramipril on ambulatory and office blood pressures: a HOPE substudy. Hypertension 2001 Dec 1;38(6):E28-E32. [doi: 10.1161/hy1101.099502] [Medline: 11751742]
46. Pickering TG, Shimbo D, Haas D. Ambulatory blood-pressure monitoring. N Engl J Med 2006 Jun 1;354(22):2368-2374. [doi: 10.1056/NEJMra060433] [Medline: 16738273]
47. O'Brien E, O'Malley K, Cox J, Stanton A. Ambulatory blood pressure monitoring in the evaluation of drug efficacy. Am Heart J 1991 Mar;121(3 Pt 2):999-1006. [doi: 10.1016/0002-8703(91)90611-K] [Medline: 1996533]
48. Peeters F, Berkhof J, Delespaul P, Rottenberg J, Nicolson NA. Diurnal mood variation in major depressive disorder. Emotion 2006 Aug;6(3):383-391. [doi: 10.1037/1528-3542.6.3.383] [Medline: 16938080]
49. Kharasch ED, Neiner A, Kraus K, Blood J, Stevens A, Schweiger J, et al. Bioequivalence and therapeutic equivalence of generic and brand bupropion in adults with major depression: a randomized clinical trial. Clin Pharmacol Ther 2018 Nov 21. [doi: 10.1002/cpt.1309] [Medline: 30460996]

50. Ekkekakis P. The Measurement of Affect, Mood, and Emotion: A Guide for Health-Behavioral Research. First Edition. New York, NY: Cambridge University Press; 2013.
51. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34 Suppl:1220-1228. [doi: 10.1037/hea0000305] [Medline: 26651463]
52. Zimmermann J, Woods W, Ritter S, Happel M, Masuhr O, Jaeger U, et al. Integrating structure and dynamics in personality assessment: first steps toward the development and validation of a personality dynamics diary. Psychol Assess 2019 Apr;31(4):516-531 [FREE Full text] [doi: 10.1037/pas0000625] [Medline: 30869961]
53. Russell JA, Carroll JM. On the bipolarity of positive and negative affect. Psychol Bull 1999 Jan;125(1):3-30. [doi: 10.1037/0033-2909.125.1.3] [Medline: 9990843]
54. Nunnally JD, Bernstein IH. Psychometric Theory. New York: Mcgraw-Hill; 1994.
55. Borsboom D. Measuring The Mind: Conceptual Issues In Contemporary Psychometrics. New York, NY: Cambridge University Press; 2019.
56. Embretson SE, Steven PR. Item Response Theory. Mahwah, NJ: CRC Press; 2013.
57. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. Evid Based Ment Health 2016 Aug;19(3):86-89 [FREE Full text] [doi: 10.1136/ebmental-2016-102418] [Medline: 27443678]
58. Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. Pers Soc Psychol Bull 2005 Oct;31(10):1340-1346. [doi: 10.1177/0146167205275615] [Medline: 16143666]
59. Parkinson B, Briner R, Reynolds S, Totterdell P. Time frames for mood: relations between momentary and generalized ratings of affect. Pers Soc Psychol Bull 2016 Jul 2;21(4):331-339. [doi: 10.1177/0146167295214003]
60. van Rijsbergen GD, Burger H, Hollon SD, Elgersma HJ, Kok GD, Dekker J, et al. How do you feel? Detection of recurrent Major Depressive Disorder using a single-item screening tool. Psychiatry Res 2014 Dec 15;220(1-2):287-293. [doi: 10.1016/j.psychres.2014.06.052] [Medline: 25070177]
61. Barge-Schaapveld DQ, Nicolson NA. Effects of antidepressant treatment on the quality of daily life: an experience sampling study. J Clin Psychiatry 2002 Jun;63(6):477-485. [doi: 10.4088/JCP.v63n0603] [Medline: 12088158]
62. Wichers MC, Barge-Schaapveld DQ, Nicolson NA, Peeters F, de Vries M, Mengelers R, et al. Reduced stress-sensitivity or increased reward experience: the psychological mechanism of response to antidepressant medication. Neuropsychopharmacology 2009 Mar;34(4):923-931 [FREE Full text] [doi: 10.1038/npp.2008.66] [Medline: 18496519]
63. Lenderking WR, Hu M, Tennen H, Cappelleri JC, Petrie CD, Rush AJ. Daily process methodology for measuring earlier antidepressant response. Contemp Clin Trials 2008 Nov;29(6):867-877. [doi: 10.1016/j.cct.2008.05.012] [Medline: 18606249]
64. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123. [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
65. Depp CA, Moore RC, Dev SI, Mausbach BT, Eyler LT, Granholm EL. The temporal course and clinical correlates of subjective impulsivity in bipolar disorder as revealed through ecological momentary assessment. J Affect Disord 2016 Mar 15;193:145-150 [FREE Full text] [doi: 10.1016/j.jad.2015.12.016] [Medline: 26773907]
66. Peters GY, Crutzen R. Pragmatic nihilism: how a Theory of Nothing can help health psychology progress. Health Psychol Rev 2017 Dec;11(2):103-121. [doi: 10.1080/17437199.2017.1284015] [Medline: 28110627]
67. Fried E, Boschloo L, van Borkulo CD, Schoevers R, Romeijn J, Wichers M, et al. Commentary: "Consistent Superiority of Selective Serotonin Reuptake Inhibitors Over Placebo in Reducing Depressed Mood in Patients with Major Depression". Front Psychiatry 2015;6:117 [FREE Full text] [doi: 10.3389/fpsyt.2015.00117] [Medline: 26347663]
68. Schultzberg M, Muthén B. Number of subjects and time points needed for multilevel time-series analysis: a simulation study of dynamic structural equation modeling. Struct Equ Modeling 2018;25:495. [doi: 10.1080/10705511.2017.1392862]
69. Trull TJ, Lane SP, Koval P, Ebner-Priemer UW. Affective dynamics in psychopathology. Emot Rev 2015 Oct;7(4):355-361 [FREE Full text] [doi: 10.1177/1754073915590617] [Medline: 27617032]
70. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
71. Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, et al. A network approach to psychopathology: new insights into clinical longitudinal data. PLoS One 2013;8(4):e60188 [FREE Full text] [doi: 10.1371/journal.pone.0060188] [Medline: 23593171]
72. Hamaker EL, Asparouhov T, Brose A, Schmiedek F, Muthén B. At the frontiers of modeling intensive longitudinal data: dynamic structural equation models for the affective measurements from the COGITO study. Multivariate Behav Res 2018 Apr 6:1-22. [doi: 10.1080/00273171.2018.1446819] [Medline: 29624092]
73. Insel T. National Institute of Mental Health. 2013. NIMH's new focus in clinical trials URL: https://tinyurl.com/y65toyyf [accessed 2018-08-07] [WebCite Cache ID 71U6GLinx]

74. Torous J, Chan SR, Yee-Marie TS, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
75. Tarver M. US Food and Drug Administration. Development of validated instruments URL: https://tinyurl.com/yy62lghz [WebCite Cache ID 71U6aIPLM]
76. US Food and Drug Administration. 1997. Guidance for industry URL: https://tinyurl.com/y28q9x5v [accessed 2018-08-07] [WebCite Cache ID 71U6cq3O8]
77. Wayback Machine. 2016. Description of the HAMD and the MADRS URL: https://tinyurl.com/yyr65kxm [accessed 2018-08-07] [WebCite Cache ID 71U6Y0K7q]
78. Lupien SJ, Sasseville M, François N, Giguère CE, Boissonneault J, Plusquellec P, Signature Consortium. The DSM5/RDoC debate on the future of mental health research: implication for studies on human stress and presentation of the signature bank. Stress 2017 Dec;20(1):95-111. [doi: 10.1080/10253890.2017.1286324] [Medline: 28124571]
79. Munos B, Baker PC, Bot BM, Crouthamel M, de Vries G, Ferguson I, et al. Mobile health: the power of wearables, sensors, and apps to transform clinical trials. Ann N Y Acad Sci 2016 Jul;1375(1):3-18. [doi: 10.1111/nyas.13117] [Medline: 27384501]

Abbreviations

CNS: central nervous system
EMA: ecological momentary assessment
EMI: ecological momentary intervention
FDA: Food and Drug Administration
HAM-D: Hamilton Rating Scale for Depression
HAM-D-17: 17-item Hamilton Rating Scale for Depression
HAM-D-24: 24-item Hamilton Rating Scale for Depression
MADRS: Montgomery Asberg Depression Scale
MBSR: mindfulness-based stress reduction
NNT: needed to treat
PRO: patient-reported outcome

Edited by J Prescott; submitted 07.08.18; peer-reviewed by J Quiroz, D Kreindler, U Ebner-Priemer, M Brandon; comments to author 19.09.18; revised version received 05.02.19; accepted 03.04.19; published 21.04.19.

Please cite as:
Mofsen AM, Rodebaugh TL, Nicol GE, Depp CA, Miller JP, Lenze EJ
When All Else Fails, Listen to the Patient: A Viewpoint on the Use of Ecological Momentary Assessment in Clinical Trials
JMIR Ment Health 2019;6(5):e11845
URL: https://mental.jmir.org/2019/5/e11845/
doi:10.2196/11845
PMID:31066701

Original Paper

Influence of Personality and Differences in Stress Processing Among Finnish Students on Interest to Use a Mobile Stress Management App: Survey Study

Mari Ervasti[1*], DSc; Johanna Kallio[1*], MSc; Ilmari Määttänen[2*], PhD; Jani Mäntyjärvi[1], DSc; Markus Jokela[2], PhD

[1]VTT Technical Research Centre of Finland Ltd, Oulu, Finland

[2]University of Helsinki, Helsinki, Finland

[*]these authors contributed equally

Corresponding Author:
Johanna Kallio, MSc
VTT Technical Research Centre of Finland Ltd
Kaitoväylä 1
Oulu, FI-90571
Finland
Phone: 358 207222434
Email: johanna.kallio@vtt.fi

Abstract

Background: Excessive stress has a negative impact on many aspects of life for both individuals and societies, from studying and working to health and well-being. Each individual has their unique level of stress-proneness, and positive or negative outcomes of stress may be affected by it. Technology-aided interventions have potential efficacy in the self-management of stress. However, current Web-based or mobile stress management solutions may not reach the individuals that would need them the most, that is, stress-sensitive people.

Objective: The aim of this study was to examine how personality is associated with stress among Finnish university students and their interest to use apps that help in managing stress.

Methods: We used 2 structured online questionnaires (combined, n=1001) that were advertised in the University of Helsinki’s mailing lists. The first questionnaire (n=635) was used to investigate intercorrelations between the Big Five personality variables (neuroticism, extraversion, openness, agreeableness, and conscientiousness) and other stress-related background variables. The second questionnaire (n=366) was used to study intercorrelations between the above-mentioned study variables and interest in using stress management apps.

Results: The quantitative findings of the first questionnaire showed that higher levels of extraversion, agreeableness, and conscientiousness were associated with lower self-reported stress. Neuroticism, in turn, was found to be strongly associated with rumination, anxiety, and depression. The findings of the second questionnaire indicated that individuals characterized by the Big Five personality traits of neuroticism and agreeableness were particularly interested to use stress management apps (r=.27, $P<.001$ and r=.11, $P=.032$, respectively). Moreover, the binary logistic regression analysis revealed that when a person’s neuroticism is one SD above average (ie, it is higher than among 84% of people), the person has roughly 2 times higher odds of being interested in using a stress management app. Respectively, when a person’s agreeableness is one SD above average, the person has almost 1.4 times higher odds of being interested in using a stress management app.

Conclusions: Our results indicated that personality traits may have an influence on the adoption interest of stress management apps. Individuals with high neuroticism are, according to our results, adaptive in the sense that they are interested in using stress management apps that may benefit them. On the contrary, low agreeableness may lead to lower interest to use the mobile stress management apps. The practical implication is that future mobile stress interventions should meaningfully be adjusted to improve user engagement and support health even among less-motivated users, for instance, to successfully engage individuals with low agreeableness.

(JMIR Ment Health 2019;6(5):e10039) doi:10.2196/10039

KEYWORDS

mental health; mobile applications; psychological stress; personality; neuroticism; mobile phone; surveys and questionnaires

Introduction

Background

The word *stress* has often somewhat differing definitions if it is used in a scientific compared with a colloquial context. From a physiological perspective, it generally refers to the loss of homeostasis of an organism [1]. Stress may have negative outcomes if it lasts too long or is too strong. On the contrary, stress may be beneficial if it is experienced in moderation relative to individual stress-sensitivity (eg, personality) and contextual features. Balancing the stress faced from the environment is a vital challenge, and thus, it is not surprising that stress is a major issue for well-being, health, and even proper functioning of the individual [2].

Stress is often referred to as if it exists in the environment per se, although it is not typically an objective feature of a specific environment. Instead, stress reactivity consists of an individual’s perception of an environment and their reaction to it [3,4]. Some individuals are more prone to stress than others, thus enabling measurement of individual stress proneness via personality and other features as well as coping styles. Personality and other descriptions of thought and behavior styles do not exist in a vacuum; they often become more self-evident after we understand the real-world consequences arising from them.

Personality and Stress Proneness

Personality and temperament are defined as characteristic patterns of thoughts, feelings, and behaviors over time and across situations [5]. These traits are relatively stable across age groups, cultures, and time [6]. In addition, mood inductions have little effect on personality inventory scores [7]. Personality has been shown to affect exposure to a stressor, influencing the nature and severity of stressor experiences [5,8,9].

Possibly, the most commonly accepted model for describing personality is currently the 5-factor model called Big Five [10,11]: it consists of 5 global factors of neuroticism (N), extraversion (E), openness to experience (O), agreeableness (A), and conscientiousness (C). In general, high scores of neuroticism are linked with emotional instability and anxiousness; high scores of extraversion are linked with socializing and impulsiveness; high scores of openness are linked with creativeness and artistry; high scores of agreeableness are linked with trustworthiness and co-cooperativeness; and high scores of conscientiousness are linked with being hardworking and vigilance [11].

Furthermore, people differ greatly in their stress proneness and coping styles. Lazarus and Folkman [12] have argued that the major determinant of coping responses is the perception of the situation by the individual. Coping has been described as “personality in action under stress” [13], and people have been shown to develop habitual methods of dealing with stressors [12]. The process of coping is seen to consist of efforts made to manage a stressful situation so that it becomes less stressful [12], and *coping strategies* are seen as categories of behavior in response to stressful events [14]. All Big Five traits have been shown to be associated with specific coping strategies [5]. In recent years, it has increasingly been proposed that different coping strategies actually overlap with personality dimensions, and thus, they should not be treated separately but rather as trait complexes. Furthermore, personality may influence the ability to implement the chosen strategy and its degree of effectiveness [5,8,9].

In general, the personality trait of neuroticism and rumination is known to be associated with negative, stress-related conditions such as depression and anxiety [15,16]. Of the key features of depressive symptoms and stress states resembling depression, 1 is often related to repetitive negative thinking. This constant rethinking of adverse events or thoughts is called rumination [17]. All people experience traumatic and stressful experiences in their lives, but some people will fixate on bad memories, thinking about them continually, thus creating a constant internal stress for themselves. Thinking about the stressful experience might seem like a way to reduce stress and anxiety, but in fact, ruminating tends to increase rumination itself.

The importance and influence of personality traits over an individual’s stress proneness, coping processes, health, and health behavior have been confirmed in several previous studies [18-20]. For instance, an earlier study [21] regarding the associations between personality and key health behaviors indicated that highly conscientious individuals had more positive health behavior, such as exercising. People with high levels of extraversion, on the contrary, were associated with risky health behavior, such as cigarette smoking and alcohol drinking. The personality trait of conscientiousness is established to be associated with outcomes that are dependent on persistently working on something over the long run, eg, education [22]. Ferguson [23] provided a comprehensive review and a theoretical model to better understand interrelationships of personality, coping, health, and health behavior.

Tailoring to Increase Intervention Effectiveness

Psychological interventions are typically resource-intensive and thus expensive. Technological approaches for delivering psychological interventions may increase their access and sustainability by being, among other things, less costly. They may extend the effective time beyond a therapy session or even replace some of the other treatment options [24]. Hence, the potential of technology tools in self-management of stress seems considerable owing to their cost-effectiveness and scalability [25].

Especially, mobile phones hold a great promise as an intervention delivery tool because they enable deeper integration into users’ everyday lives [26]. Thus, it seems feasible to implement health interventions using mobile technology; participants have demonstrated adequate compliance with the treatment protocol and have described momentary interventions as a credible and acceptable form of treatment [27]. Earlier researches have already noted that mental health interventions delivered through mobile apps can be effective in treating a range of mental health problems, such as depression, stress, and anxiety [28]. The use of mobile apps has been shown to enhance the sense of comfort and acceptance of the intervention [29,30].

However, mobile intervention solutions have not sufficiently taken into account the differences between individual users. Tailoring refers to the system autonomously performing the tailoring of personally relevant content without any direct user intervention through applying a preselected self-management strategy [27]. Technology-based systems can automatically tailor intervention content based on characteristics or information provided by individuals during preintervention or momentary assessments. Tailoring and smooth interaction with daily life appear to be more important for long-term adherence and beneficial outcomes than a multitude of options [31,32]. It has been shown that tailoring the content of health messages based on individual characteristics can improve message acceptance and willingness to change [33,34]. Furthermore, personally relevant content and tools may lead to better adherence and more lasting effects, especially in preventive programs [35,36].

Research Questions

Personality influences people's decision-making processes, and information about personality can be utilized in designing health apps [30]. We selected study variables known to be associated with stress reactivity, everyday emotions, and outcomes of long-term stress. Our online questionnaires included questions about personality, other processing styles of everyday information (rumination, self-reflection, and self-directedness), depressive symptoms, anxiety, coping styles, health status, and interest in using stress management apps. Relevant personal information about coping, personality, and related variables could be useful in tailoring individual interventions for better self-management of stress [37].

Personality and other traits involving stress reactivity and stress proneness may be useful in understanding the following empirical questions that are relevant for interventions, especially mobile interventions: (1) who are the individuals reporting most health and mental health problems, (2) who are most likely to start participating in an intervention, (3) who are most likely to continue participating in an intervention, (4) who are the ones benefitting the most from interventions, and (5) how could feedback and other features be tailored according to individual differences. The research variables and research questions of the study were designed to answer, when suitable, as many of these questions as possible.

With this exploratory study, we were interested in finding answers to the following questions:

1. How are an individual's personality and stress-related background variables associated with (1) each other and (2) health and mental health variables?
2. How are the individual's personality and stress-related background variables linked to the interest to use stress management apps?

Methods

Data Collection

We approached our research questions in an exploratory way, starting from the existing research and theories and then examining how people's personality is related to stress and to their interest to use apps that help in stress management. Data were collected via 2 online questionnaires. The study was performed in accordance with ethical guidelines and it was reviewed by the Ethics Review Board in humanities and social and behavior sciences of the University of Helsinki. Table 1 presents a summary of the data.

The Big Five personality variables were assessed using the Neuroticism, Extraversion, Openness-Five-Factor Inventory (NEO-FFI) [38]. NEO-FFI is a shorter version of the Revised NEO Personality Inventory, and it contains 5 subsections that assess the factors of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. There are 12 items per subsection, for a total of 60 items. Participants were asked to indicate their response to each item on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). Higher scores indicated higher levels of each personality dimension.

Self-reported stress was studied with a single item stress question and a description of stress (translation: "Stress means a situation in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because his/her mind is troubled all the time. Do you feel this kind of stress these days?") [39]. Depressive symptoms were assessed using the Center for Epidemiologic Studies Depression Scale questionnaire [40], and anxiety was assessed using the Generalized Anxiety Disorder questionnaire [41]. Rumination tendency was explored through the questionnaire devised by Elliot and Coker [42].

Moreover, the online questionnaire included basic queries about background information (eg, gender, age, education, work, social, and health status). The following questions were asked regarding the perceived usage activity of mobile apps and interest in using novel mobile apps: "Do you have a mobile phone?," "Do you consider yourself as an active user of mobile application?," and "Are you interested in using novel mobile applications on a daily basis?"

Table 1. Summary of data collection methods, numbers of participants, and goals of studies 1 and 2.

Data collection method	Participants	Goal of study
Study 1: structured online questionnaire	635 university students	To gain knowledge of the links between personality traits and stress-related mental disturbances such as depressive symptoms, anxiety, and rumination
Study 2: structured online questionnaire	366 university students	To explore how people's interests for usage of technology-aided stress management solutions are related to their personality, stress proneness, depressive symptoms, and rumination

The second online questionnaire was otherwise similar, but it was somewhat improved by adding a few questions based on insights gained through the first survey round. The questions added to Study 2 were the following: “What kinds of stress management or relieving methods you are using?”, “Do you have previous experience on well-being applications?”, “If yes, what kind of application you have used?”, and “Are you interested in using a self-management application for managing stress?”

Apart from the previously mentioned items and questionnaire, Siegrist’s work stress measure (effort-reward imbalance, ERI) [43] was included in the questionnaires but was not used in this study. The sample consisted of mostly full-time students, the resulting ERI was less suitable for the sample. In addition to having few people actually reporting their work stress, the persons who reported the ERI were not likely full-time workers in professions that would match their educational background. Thus, the ERI was deemed as not suitable for this study.

The quantitative analysis of the questionnaire data was performed with IBM SPSS Statistics software. The Pearson correlation coefficient was used in the analyses.

Study 1: First Online Questionnaire

The structured online questionnaire was used with the purpose of reaching a large sample of respondents representing a fairly heterogeneous group of individuals. Participants (n=635) were voluntarily recruited from the University of Helsinki, Finland, by sending an invitation via email to several of the distribution lists of university students in November 2014. The invitation contained a link to an electronic questionnaire, which was provided by the University’s Web service, *E-lomake*. Over 99% of the respondents responded within 14 days of the questionnaire being accessed.

Most of the respondents were female (86.9%; 552/635) and full-time students (79.5%; 505/635). Most of the respondents were 26 years or younger (69.0%; 438/635), and the age range was 18 to 62 years. All respondents responded in Finnish. It may be useful to clarify that there are practically no ethnic minorities among the Finnish-speaking students in the University of Helsinki; apart from Swedish-speaking Finns and potentially adopted children, there were no minorities in the sample or their number was extremely small (<1%). Thus, no further information regarding ethnicity is provided.

The socioeconomic background of the sample was relatively evenly spread; 25.4% (161/635) reported their family background as *working class*, 49.0% (311/635) as *middle class*, 23.6% (150/635) as *upper middle class*, and 1.9% (12/635) as *upper class*. Of the whole sample, 23.3% (148/635) had had a diagnosis of clinical depression at some point in their lives and 7.2% (46/635) were currently clinically depressed. Altogether, 14.3% (91/635) of the respondents reported having worse or much worse health than average people of their age, whereas 33.4% (212/635) reported having better or much better health than the average person of their age. Moreover, 22.5% (143/635) of respondents reported having difficulties *quite often* or *often* in making ends meet (=not having enough money). Of the respondents, 14.0% (89/635) reported being *rather dissatisfied* or *very dissatisfied* with their lives.

More than half of the respondents considered themselves as active mobile app users (56.1%; 356/635), and nearly half were interested in using novel mobile apps on a daily basis (49.1%; 312/635).

Study 2: Second Online Questionnaire

The second online questionnaire (n=366) was similar to the Study 1 questionnaire except for the 4 supplementary questions concerning stress management and experience of well-being apps as presented in the Data Collection section. The data collection and analysis methods were similar to the first study. Email invitations were sent in February 2015. Over 97% of the respondents responded within 14 days of opening the questionnaire.

Most of the respondents were female (83.9%; 308/366) and full-time students (74.8%; 274/366). The age range was 19 to 56 years, with the majority of respondents being 30 years or younger (n=304; 83.0%; 304/366). Slightly over half of the respondents considered themselves as active mobile app users (57.1%; 209/366), and nearly half reported being interested in using novel mobile apps on a daily basis (45.9%; 168/366). Ethnicity information is similar to that in Study 1; either there were basically no ethnic minorities in the sample or their number was extremely small (<1%).

As with Study 1, the socioeconomic background of the sample was relatively evenly spread; 24.9% (91/366) reported their family background as *working class*, 46.7% (171/366) as *middle class*, 26.8% (98/366) as *upper middle class*, and 1.6% (6/366) as *upper class*. Of the whole sample, 23.0% (84/366) had had a diagnosis of clinical depression at some point in their lives and 5.4% were currently clinically depressed. In addition, 17.5% (64/366) of the respondents reported worse or much worse health than average people of their age, whereas 35.6% (130/366) reported better or much better health than the average person of their age. Altogether, 23.0% (84/366) of respondents reported having difficulties *quite often* or *often* in making ends meet (=not having enough money). Moreover, 12.3% (45/366) of respondents described being *rather dissatisfied* or *very dissatisfied* with their lives.

Every fifth respondent (20.8%; 76/366) had previous experience with well-being apps, and 55.5% (203/366) stated that they would be interested in using apps geared to stress management.

Results

Findings From Study 1

In the first questionnaire round, we investigated correlations between the Big Five personality dimensions and other stress-related variables. The correlations are presented in Multimedia Appendix 1.

Neuroticism was strongly associated with rumination, anxiety, and depressive symptoms. People with higher neuroticism reported more stress and were more likely to have a diagnosis of clinical depression. Furthermore, higher neuroticism was associated with lower happiness and being less content in the current life situation. Higher neuroticism was also associated with a history of absences from work or similar duties for psychological reasons and worry about the financial situation and was negatively associated with social status in society and personal health situation.

Extraversion had a strong negative association with anxiety, depressive symptoms, and rumination. Furthermore, extraversion seemed to be associated with a higher level of happiness in the current life situation and social status together with the prevailing health situation.

People characterized by the personality trait of agreeableness were discovered to have a lower tendency for rumination, self-reported stress, depressive symptoms, and anxiety. This personality trait appeared to also be associated with general happiness in life as well as with satisfaction with current social status, health, and financial situation.

Conscientiousness had a positive association with happiness in life, social status, and health situation. However, it was negatively associated with the perceived financial situation. This personality trait was negatively associated with rumination, anxiety, depressive symptoms, and self-reported stress.

The personality trait of openness to experience was positively associated with satisfaction with one's life and financial situation, and interestingly, people characterized by openness appeared to be more prone to self-reported stress and absences from work owing to psychological reasons.

Findings From Study 2

We investigated how a person's characteristics correspond to actual interest to use stress management apps and whether it would be possible to distinguish a potential user population with specific personality traits and stress proneness. This was initially evaluated by correlating questionnaire variables with the survey question "Are you interested in using applications that help in stress management?;" significant correlations are presented in Table 2.

The correlation analysis revealed that previous stressful life events, current self-reported stress, diagnosed depression, history of absences from work for psychological reasons, and previous experience of psychological treatment were significantly related to interest for stress management app use. Thus, individuals who are stressed and depressed appear to be the most in need of the app. Furthermore, current health situation, financial situation, and perceived general happiness in the current life situation were correlated with stress management app usage interest. Stress proneness and perceived past stressful life events and current self-reported stress therefore appear to be associated with higher levels of acceptance of a stress management solution and with the interest to use the technology.

Rumination, anxiety, and depressive symptoms were significantly associated with usage interest. Individuals characterized by the Big Five personality traits of neuroticism and agreeableness appeared to have a significant positive association with interest to use stress management apps, with r=.27 and .11, respectively. The first questionnaire round already identified the strong correlation between neuroticism and the tendency for rumination and depressive symptoms, among others.

We also examined whether respondents' general interest in using novel mobile apps on a daily basis and previous experience of using well-being apps correlated with personality and stress-related background variables. Interestingly, no significant correlations emerged between these variables.

Owing to associations with neuroticism and other personality traits, the associations between the coping styles, rumination, and other variables shown in Table 2, and the dependent variables, were not statistically significant in the combined model of logistic regression analysis (*P*>.05). The conducted analysis did not include coping styles and rumination for said reasons. The binary logistic regression analysis included standardized values for covariates of personality traits (N, E, A, C, and O), sex and age. The regression analysis showed statistically significant associations between the interest to use stress management apps and neuroticism (*P*<.001) and agreeableness (*P*=.010). When a person's neuroticism is one SD above the mean, that is, larger than among 68% of the people, he or she has 2.036 times higher odds (95% CI 1.543 to 2.687) to be interested in using stress management apps. Respectively, when a person's agreeableness is higher than the mean by one SD, the person has 1.379 higher odds to be interested in using stress management apps (95% CI 1.081 to 1.758). Sex did not have a significant association with the interest in using stress management apps, when analysed together with the said personality traits and age.

Table 2. Intercorrelations between Study 2 variables and interest to use stress management apps. Significant associations are indicated in italics.

Variable	Interest in using novel mobile apps on a daily basis		Interest in using stress management app	
	r value	*P* value	r value	*P* value
Age	−.05	.35	−.02	.78
Happiness	*− .11*	*.03*	*− .11*	*.04*
Social status	−.01	.86	−.05	.38
Financial situation	*.11*	*.03*	*.15*	*.003*
Health situation	−.06	.29	*− .16*	*.002*
Rumination	.03	.63	*.19*	*<.001*
Self-reported stress	.02	.73	*.24*	*<.001*
Stressful and hard life-event within past 5 years	.06	.22	*.16*	*.002*
Reaction to this life-event with a long-term reduction of performance	−.04	.43	*− .13*	*.01*
Absence from work for psychological reasons	.04	.44	*.17*	*<.001*
Professional help for psychological problems	.03	.56	*.18*	*<.001*
Depressive symptoms	.07	.16	*.19*	*<.001*
Anxiety	.08	.14	*.18*	*<.001*
Neuroticism	.05	.38	*.27*	*<.001*
Extraversion	.08	.15	.01	.90
Agreeableness	.00	>.99	*.11*	*.03*
Conscientiousness	*− .11*	*.04*	−.09	.09
Openness to experience	.08	.15	.01	.80

Discussion

Research Question 1: Personality's Influence on Stress Proneness and Coping (Study 1)

The findings of the first questionnaire revealed that, as expected, neuroticism was strongly linked to rumination, anxiety, and depressive symptoms. Furthermore, people characterized by neuroticism experienced more self-reported stress and reported being less happy and content in their current life situation, consistent with previous research [5]. Thus, the study population was similar to other populations studied with the said variables.

Extraversion may act as a protective factor for stress-related disorders because it was associated negatively with anxiety, depressive symptoms, and rumination. This finding is also congruent with the results of Takano and Tanno [16] and Bunevicius et al [22], who showed that higher levels of extraversion are negatively associated with stress. Agreeableness was found to diminish the tendency for rumination, self-reported stress, depressive symptoms, and anxiety. Conscientiousness was associated positively with happiness in life, social status, and health situation. In addition, conscientiousness associated negatively with rumination, anxiety, depressive symptoms, and self-reported stress. Openness to experience was positively associated with satisfaction with one's financial situation and, perhaps surprisingly, with self-reported stress and absence from work for psychological reasons. Overall, higher levels of extraversion, agreeableness, and conscientiousness were associated with lower self-reported stress, thus providing possible benefits for stress outcomes.

Research Question 2: Stress Management App Usage Interests (Study 2)

After the initial correlation analysis, we conducted a binary logistic regression analysis with standardized variables. The results of the analysis suggested that individuals characterized by the Big Five personality traits of neuroticism and agreeableness are more likely interested to use stress management apps. People with higher neuroticism (or lower emotional stability), being more stress-sensitive, seem to be adaptive in this regard. When it comes to interest in stress management apps, their personal topics of interest coincide with their own benefit. On the contrary, higher agreeableness could increase the odds of being interested in using stress intervention software. Odds ratio between neuroticism and usage interest indicated that increase by one SD of neuroticism is associated with 2.036 times higher odds to be interested in using stress management apps. Respectively, the odds ratio between usage interest and agreeableness was 1.379. This is a particularly interesting finding as neuroticism and agreeableness were found to be negatively associated ($r=-.24$, $P<.001$). Thus, there are likely to be subgroups of people who would benefit from a stress management app but who are inhibited by being interested in using them by some other trait, such as a very low agreeableness.

The findings support earlier research showing that personality traits are important contributors to technology acceptance [44].

For instance, the tolerant and cooperative nature of agreeable people makes them easier to accept new technologies, in general [45]. However, the possible impact of personality traits on the adoption of mobile apps has been shown to differ between different app categories [46]. Our findings suggest that especially individuals with high agreeableness and high neuroticism are more likely to be interested in using mobile stress management solutions.

Earlier research has emphasized the importance of gaining deep insights into a person's personality, stress proneness, and coping strategies [14]. Another consideration should be to avoid burdening the users with data collection activities, instead using the data already collected [47]. Previous studies have shown that people may avoid attending face-to-face therapy as mental health problems and illnesses are still a taboo for many people. In the future, mobile stress self-management tools could be used in combination with traditional care as people who suffer from depressive symptoms or related conditions could benefit from external motivation to encourage themselves to use self-management solutions for relieving their mental symptoms [48]. In addition, people's attitudes toward mobile phones as monitoring and self-management tools for mental health have been shown to be positive [49,50].

The findings suggest that tailoring the content of interventions according to personality and personal needs is essential to make digital interventions more personally relevant. Greater practitioner and end-user involvement in the co-design process would help avoid a mismatch between technology and the designated context of use [51,52]. Tailoring should not be done on a superficial level but should take into account the user's personal situation and psychological needs. Tailored messages have proven to improve intervention acceptance and efficacy, thus increasing the willingness to change and boosting treatment adherence and engagement rates [34,53].

Limitations and Future Work

There are some noteworthy limitations in this study. As an exploratory study, the findings cannot be generalized too widely. Our respondents do not reflect the average population because the majority of the questionnaire respondents were full-time students and women aged under 30 years. To further complicate our study on the feasibility and validity of profiling survey respondents, personality appears to be more strongly associated with coping in young samples [5]. Stress coping strategies seem to change over time as responses to stress are driven more strongly by temperament in younger individuals [5]. Gender may also moderate relations between personality and coping styles owing to sex differences in the types of stressors experienced [54]. Another important aspect to consider is that the study was conducted among Finnish people. As attitudes toward stress and personality variables may differ between cultures, the findings may not be directly transferrable to other cultural settings. It is possible that the characteristics of our respondent population influenced the results, as they are by no means representative of all adults. Our results should be further investigated and confirmed in more diverse populations. In addition, multiple correlations were analyzed and presented simultaneously in the results, which might itself be somewhat problematic. On the contrary, the significances of correlations were relatively high and the results were logical and in line with each other and with previous findings, with no large outliers. Thus, the results may be considered relatively reliable in that sense.

Our future work will focus on developing a stress management app that supports users in their everyday life with personally meaningful interventions. It would be beneficial to highlight and create more personalized solutions for the groups that could benefit from stress-reduction interventions but who have other traits that make their interest less likely, such as people at the higher end of neuroticism and very low agreeableness. According to previous research, individuals characterized by different personality dimensions are accustomed to using different coping strategies. We hypothesized that after having identified the user's personality, we could define distinct user profiles and identify preferred coping strategies and intervention(s). Thus, we plan to merge user profiles and characteristic coping styles to validate the approach of automatically determining and tailoring the content of mobile interventions.

Conclusions

This study serves as a basis for a mobile service development process helping to derive design implications for a stress management solution that would be effective and acceptable. We studied how personality is associated with stress among Finnish university students and how personality may influence on an individual's interest to use a self-management app for managing stress. According to our findings, individuals characterized by the personality traits of neuroticism and agreeableness are more likely to be interested to use stress management apps. These findings suggest that stress management apps should consider tailoring content based on the user's personality or similar constructs to enhance user adherence and engagement. However, further research on how to best accomplish this is needed.

Acknowledgments

This work was supported by the Innovation Funding Agency Tekes (now Business Finland). In addition, the authors want to acknowledge the valuable contribution of the people who participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Intercorrelations (r value; P value) between Study 1 variables.

[PDF File (Adobe PDF File), 25KB - mental_v6i5e10039_app1.pdf]

References

1. Lazarus RS. From psychological stress to the emotions: a history of changing outlooks. Annu Rev Psychol 1993;44:1-21. [doi: 10.1146/annurev.ps.44.020193.000245] [Medline: 8434890]
2. European Agency for Safety and Health at Work. Luxembourg: Publications Office of the European Union; 2017. Healthy Workplaces for All Ages 2016-2017 URL: https://tinyurl.com/y2luh5y5 [accessed 2019-02-17] [WebCite Cache ID 76Gg49K9M]
3. Ortiz MS, Willey JF, Chiang JJ. [How stress gets under the skin]. Rev Med Chil 2014 Jun;142(6):767-774 [FREE Full text] [doi: 10.4067/S0034-98872014000600011] [Medline: 25327322]
4. Määttänen I, Ravaja N, Henttonen P, Puttonen S, Paavonen K, Swan H, et al. Type 1 long QT syndrome and psychological stress in a laboratory setting. J Health Psychol 2018 Jan 1:1359105317751617. [doi: 10.1177/1359105317751617] [Medline: 29355047]
5. Connor-Smith JK, Flachsbart C. Relations between personality and coping: a meta-analysis. J Pers Soc Psychol 2007 Dec;93(6):1080-1107. [doi: 10.1037/0022-3514.93.6.1080] [Medline: 18072856]
6. McCrae RR, Costa PT, Del Pilar GH, Rolland J, Parker WD. Cross-cultural assessment of the five-factor model. J Cross Cult Psychol 2016 Jul 27;29(1):171-188. [doi: 10.1177/0022022198291009]
7. McCrae R, Costa T. Emerging Lives, Enduring Dispositions: Personality in Adulthood. Boston: Little, Brown and Co; 1984.
8. Bolger N, Zuckerman A. A framework for studying personality in the stress process. J Pers Soc Psychol 1995 Nov;69(5):890-902 [FREE Full text] [doi: 10.1037//0022-3514.69.5.890] [Medline: 7473036]
9. Burgess L, Irvine F, Wallymahmed A. Personality, stress and coping in intensive care nurses: a descriptive exploratory study. Nurs Crit Care 2010;15(3):129-140. [doi: 10.1111/j.1478-5153.2009.00384.x] [Medline: 20500651]
10. John O, Srivastava S. The Big Five trait taxonomy: history, measurement,theoretical perspectives. In: Pervin AL, John OP, editors. Handbook of Personality:Theory and Research. New York: Guilford; 1999:102-138.
11. Costa PT, McCrae RR. Revised NEO Personality Inventory (NEO PI-R). Odessa: Psychological Assessment Resources; 1992.
12. Lazarus R, Folkman S. Stress, Appraisal, And Coping. New York: Springer Publishing Company; 1984.
13. Bolger N. Coping as a personality process: a prospective study. J Pers Soc Psychol 1990 Sep;59(3):525-537 [FREE Full text] [doi: 10.1037//0022-3514.59.3.525] [Medline: 2231283]
14. McCrae RR, Costa PT. Personality, coping, and coping effectiveness in an adult sample. J Personality 1986 Jun;54(2):385-404. [doi: 10.1111/j.1467-6494.1986.tb00401.x]
15. Kotov R, Gamez W, Schmidt F, Watson D. Linking “big” personality personality traits to anxiety, depressive, and substance use disorders: A meta-analysis.. Psychol Bull 2010 Sep;136(5):768-821. [doi: 10.1037/a0020327] [Medline: 20804236]
16. Takano K, Tanno Y. Self-rumination, self-reflection, and depression: self-rumination counteracts the adaptive effect of self-reflection. Behav Res Ther 2009 Mar;47(3):260-264. [doi: 10.1016/j.brat.2008.12.008] [Medline: 19181307]
17. Sekihara Y, Konno K, Kawaguchi T. [Detection of the occluded artery in cerebral infarction by MRI]. No Shinkei Geka 1989 Dec;17(12):1145-1151. [Medline: 2615898]
18. Ferguson E. Personality and coping traits: a joint factor analysis. Br J Health Psychol 2001 Nov;6(Part 4):311-325. [doi: 10.1348/135910701169232] [Medline: 12614507]
19. Turner RD, Vollmer W, Foster SJ. Different walls for rods and balls: the diversity of peptidoglycan. Mol Microbiol 2014 Mar;91(5):862-874 [FREE Full text] [doi: 10.1111/mmi.12513] [Medline: 24405365]
20. Vollrath M, Knoch D, Cassano L. Personality, risky health behaviour, and perceived susceptibility to health risks. Eur J Pers 1999;13(1):39-50 [FREE Full text] [doi: 10.1002/(SICI)1099-0984(199901/02)13:1<39::AID-PER328>3.0.CO;2-J]
21. Raynor DA, Levine H. Associations between the five-factor model of personality and health behaviors among college students. J Am Coll Health 2009;58(1):73-81. [doi: 10.3200/JACH.58.1.73-82] [Medline: 19592356]
22. Bunevicius A, Katkute A, Bunevicius R. Symptoms of anxiety and depression in medical students and in humanities students: relationship with big-five personality dimensions and vulnerability to stress. Int J Soc Psychiatry 2008 Nov;54(6):494-501. [doi: 10.1177/0020764008090843] [Medline: 18974188]
23. Ferguson E. Personality is of central concern to understand health: towards a theoretical model for health psychology. Health Psychol Rev 2013 May;7(Suppl 1):S32-S70 [FREE Full text] [doi: 10.1080/17437199.2010.547985] [Medline: 23772230]
24. Brizinskiĭ MA, Ivashkovskiĭ AI. [A case of idiopathic osteoarthropathy]. Vrach Delo 1989 Oct(10):92-94. [Medline: 2618000]

25. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
26. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
27. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
28. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
29. Ahtinen A, Mattila E, Välkkynen P, Kaipainen K, Vanhala T, Ermes M, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR Mhealth Uhealth 2013;1(2):e11 [FREE Full text] [doi: 10.2196/mhealth.2596] [Medline: 25100683]
30. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
31. Walter A. Designing For Emotion. New York: A Book Apart; 2011.
32. Ahtinen A, Mattila E, Väätänen A, Hynninen L, Salminen J, Koskinen E, et al. User experiences of mobile wellness applications in health promotion: User study of Wellness Diary, Mobile Coach and SelfRelax. : IEEE; 2009 Apr Presented at: 2009 3rd International Conference on Pervasive Computing Technologies for Healthcare; April 1-3, 2009; London p. 1-8. [doi: 10.4108/ICST.PERVASIVEHEALTH2009.6007]
33. Kreuter M, Farrell D, Olevitch L, Brennan L, editors. Tailoring Health Messages: Customizing Communication with Computer Technology. New York: Routledge Communication Series; 1999.
34. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999;21(4):276-283. [doi: 10.1007/BF02895958] [Medline: 10721433]
35. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: An exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.: ACM; 2012 May Presented at: CHI'12; May 5-10, 2012; Austin, Texas p. 1421-1430. [doi: 10.1145/2207676.2208602]
36. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
37. Halko S, Kientz J. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. 2010 Presented at: International Conference on Persuasive Technology; June 7-10, 2010; Copenhagen, Denmark p. 150-161.
38. Konstabel K, Lönnqvist J, Walkowitz G, Konstabel K, Verkasalo M. The ‘Short Five’ (S5): measuring personality traits using comprehensive single items. Eur J Pers 2011 Feb 17;26(1):13-29. [doi: 10.1002/per.813]
39. Elo A, Leppänen A, Jahkola A. Validity of a single-item measure of stress symptoms. Scand J Work Environ Health 2003 Dec;29(6):444-451 [FREE Full text] [doi: 10.5271/sjweh.752] [Medline: 14712852]
40. Radloff LS. The CES-D scale: a self report depression scale for research in the general population. Appl Psychol Meas 1977 Jun 1;1(3):385-401. [doi: 10.1177/014662167700100306]
41. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
42. Elliott I, Coker S. Independent self‐construal, self‐reflection, and self‐rumination: a path model for predicting happiness. Aust J Psychol 2008;60(3):127-134. [doi: 10.1080/00049530701447368]
43. Siegrist J. European Network for Mental Health Promotion. 2012. Effort-reward imbalance at work - theory, measurement and evidence URL: http://www.mentalhealthpromotion.net/resources/theorie_measurement_evidence.pdf [accessed 2019-02-18] [WebCite Cache ID 76GpE4hvr]
44. Svendsen GB, Johnsen JK, Almås-Sørensen L, Vittersø J. Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model. Behav Inf Technol 2013 Apr;32(4):323-334. [doi: 10.1080/0144929X.2011.553740]
45. Devaraj S, Easley RF, Crant JM. Research note: how does personality matter? Relating the five-factor model to technology acceptance and use. Info Syst Res 2008 Mar;19(1):93-105. [doi: 10.1287/isre.1070.0153]
46. Xu R, Frey RM, Fleisch E, Ilic A. Understanding the impact of personality traits on mobile app adoption – insights from a large-scale field study. Comp Hum Behav 2016 Sep;62:244-256. [doi: 10.1016/j.chb.2016.04.011]
47. Muuraiskangas S, Harjumaa M, Kaipainen K, Ermes M. Process and effects evaluation of a digital mental health intervention targeted at improving occupational well-being: lessons from an intervention study with failed adoption. JMIR Ment Health 2016 May 11;3(2):e13 [FREE Full text] [doi: 10.2196/mental.4465] [Medline: 27170553]
48. Harjumaa M, Halttu K, Koistinen K, Oinas-Kukkonen H. User Experience of Mobile Coaching for Stress-Management to Tackle Prevalent Health Complaints. : Springer; 2015 Aug Presented at: Scandinavian Conference on Information Systems; August 9-12, 2015; Oulu, Finland p. 152-164. [doi: 10.1007/978-3-319-21783-3_11]

49. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
50. Apolinário-Hagen J, Harrer M, Kählke F, Fritsche L, Salewski C, Ebert DD. Public attitudes toward guided internet-based therapies: web-based survey study. JMIR Ment Health 2018 May 15;5(2):e10735 [FREE Full text] [doi: 10.2196/10735] [Medline: 29764797]
51. Ervasti M. Understanding and predicting customer behaviour: framework of value dimensions in mobile services. J Cust Behav 2013 Nov 29;12(2):135-158. [doi: 10.1362/147539213X13832198548256]
52. van Velsen L, Wentzel J, van Gemert-Pijnen GJ. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
53. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
54. Leandro PG, Castillo MD. Coping with stress and its relationship with personality dimensions, anxiety, and depression. Procedia Soc Behav Sci 2010;5:1562-1573. [doi: 10.1016/j.sbspro.2010.07.326]

Abbreviations

ERI: effort-reward imbalance
NEO-FFI: Neuroticism, Extraversion, Openness-Five-Factor Inventory

Edited by V Bhat; submitted 05.02.18; peer-reviewed by K Kaipainen, J Fillo; comments to author 24.03.18; revised version received 29.06.18; accepted 27.01.19; published 13.05.19.

Please cite as:
Ervasti M, Kallio J, Määttänen I, Mäntyjärvi J, Jokela M
Influence of Personality and Differences in Stress Processing Among Finnish Students on Interest to Use a Mobile Stress Management App: Survey Study
JMIR Ment Health 2019;6(5):e10039
URL: https://mental.jmir.org/2019/5/e10039/
doi:10.2196/10039
PMID:31094358

Original Paper

Economic Evaluation of an Internet-Based Stress Management Intervention Alongside a Randomized Controlled Trial

Fanny Kählke[1], MPH, MEd; Claudia Buntrock[1], PhD; Filip Smit[2,3,4], PhD; Matthias Berking[1], PhD; Dirk Lehr[5], PhD; Elena Heber[6], PhD; Burkhardt Funk[7], PhD; Heleen Riper[2,8], PhD; David Daniel Ebert[1,2], PhD

[1]Institute for Psychology, Department of Clinical Psychology and Psychotherapy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[2]Amsterdam Public Health, Department of Clinical, Neuro- and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands

[3]Amsterdam Public Health Research Institute, Department of Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, Netherlands

[4]Netherlands Institute of Public Mental Health, Centre of Health-Economic Evaluation, Trimbos Institute, Utrecht, Netherlands

[5]Institute for Psychology, Department of Health Psychology and Applied Biological Psychology, Leuphana University, Lüneburg, Germany

[6]GET.ON Institute for Online Health Trainings, Hamburg, Germany

[7]Institute of Information Systems, Leuphana University, Lüneburg, Germany

[8]Telepsychiatric Centre, University of Southern Denmark, Odense, Denmark

Corresponding Author:
Fanny Kählke, MPH, MEd
Institute for Psychology
Department of Clinical Psychology and Psychotherapy
Friedrich-Alexander-Universität Erlangen-Nürnberg
2nd Floor
Nägelsbachstrasse 25a
Erlangen, 91052
Germany
Phone: 49 9131 85 ext 67568
Fax: 49 9131 85 67576
Email: fanny.kaehlke@fau.de

Abstract

Background: Work-related stress is widespread among employees and associated with high costs for German society. Internet-based stress management interventions (iSMIs) are effective in reducing such stress. However, evidence for their cost-effectiveness is scant.

Objective: The aim of this study was to assess the cost-effectiveness of a guided iSMI for employees.

Methods: A sample of 264 employees with elevated symptoms of perceived stress (Perceived Stress Scale≥22) was assigned to either the iSMI or a waitlist control condition (WLC) with unrestricted access to treatment as usual. Participants were recruited in Germany in 2013 and followed through 2014, and data were analyzed in 2017. The iSMI consisted of 7 sessions plus 1 booster session. It was based on problem-solving therapy and emotion regulation techniques. Costs were measured from the societal perspective, including all direct and indirect medical costs. We performed a cost-effectiveness analysis and a cost-utility analysis relating costs to a symptom-free person and quality-adjusted life years (QALYs) gained, respectively. Sampling uncertainty was handled using nonparametric bootstrapping (N=5000).

Results: When the society is not willing to pay anything to get an additional symptom-free person (eg, willingness-to-pay [WTP]=€0), there was a 70% probability that the intervention is more cost-effective than WLC. This probability rose to 85% and 93% when the society is willing to pay €1000 and €2000, respectively, for achieving an additional symptom-free person. The cost-utility analysis yielded a 76% probability that the intervention is more cost-effective than WLC at a conservative WTP threshold of €20,000 (US $25,800) per QALY gained.

Conclusions: Offering an iSMI to stressed employees has an acceptable likelihood of being cost-effective compared with WLC.

Trial Registration: German Clinical Trials Register DRKS00004749; https://www.drks.de/DRKS00004749

International Registered Report Identifier (IRRID): RR2-10.1186/1471-2458-13-655

(JMIR Ment Health 2019;6(5):e10866) doi:10.2196/10866

KEYWORDS

work; occupational stress; economic evaluation; internet; quality of life; clinical trials, randomized

Introduction

Background

Up to 27% of the workforce in Europe suffers from elevated stress levels [1]. According to the effort-reward imbalance model [2] and the job demand-control model [3] situations characterized by an imbalance between high effort (eg, workload) and low reward (eg, job insecurity) or high demand and low job decision latitude lead to high levels of strain. This strain is known to be a risk factor for psychological and physiological health consequences such as sleeping problems [4], mental health problems [5], cardiovascular disease [6], and chronic pain [7]. Consequently, the resulting economic burden due to productivity losses (eg, sick leave) [8] and higher health care consumption and out-of-pocket payments is substantial [9-11]. The estimated costs of work-related stress range from US $221.13 million to US $187 billion and therefore impose a tremendous burden on society [12]. Psychological interventions can be effective in reducing stress [13], but the availability of face-to-face treatments is limited [14]. Web-based and mobile-based interventions have been proposed to overcome the limitations of traditional face-to-face interventions. Such interventions are low-threshold interventions, are available 24/7, and are associated with low costs [15].

In a recent meta-analysis, it has been shown that internet-based stress management interventions (iSMIs) are effective with an effect size of *d*=0.43 (95% CI 0.31-0.51) on perceived stress [16] and a small effect on depression and anxiety, but lack evidence regarding cost-effectiveness. Internet-based interventions are often argued to be cost-effective, yet there exists little evidence. Donker et al [17] found that internet-based interventions for common mental health disorders have a considerable probability of being more cost-effective when compared with control groups. Most health economic outcome studies evaluated internet-based interventions for alcohol consumption [18], smoking cessation [19], anxiety [20], and depression [17,21]. Hedman et al compared an iSMI with an internet-based cognitive behavior therapy (iCBT) for treatment of health anxiety, where the iSMI resulted in lower costs [22].

Objectives

To our knowledge, there exist no cost-effectiveness and cost-utility analyses of iSMIs from the societal perspective. Thus our aim was to establish the cost-effectiveness and cost-utility of this iSMI for employees.

Methods

Design

This study is a health-economic evaluation with a 6-month time horizon from a societal perspective alongside a 2-arm randomized controlled trial (RCT) in Germany to establish the cost-effectiveness and cost-utility of an iSMI for employees with elevated work-related stress in combination with usual care compared with a waitlist control condition (WLC) with access to treatment as usual [23]. The present health-economic evaluation followed guidelines from the International Society for Pharmacoeconomics and Outcomes Research RCT-cost-effectiveness analysis Task Force report and the recommendations of the Consolidated Health Economic Evaluation Reporting Standard [24,25]. The trial included 264 participants who were randomly allocated in a 1:1 ratio with a block size of 2 to either iSMI or WLC. An independent researcher not otherwise involved in the study performed the randomization using randomization software (Randlist, Datinf GmbH) [26]. Participants were included in the study if they were 18 years or older, currently employed, and scored 22 or above on the Perceived Stress Scale (PSS-10). One SD (SD 6.2) above the mean (PSS-10=15.3) in a large working population [27] was chosen as a cut-off value to select participants with an elevated level of stress. The exclusion criteria were to be at risk of suicide or dissociative symptoms or having been diagnosed with a psychosis. The Ethics Committee of the Philipps-University of Marburg, Germany, approved the study. The trial was registered (DRKS00004749) in the German Clinical Trials Register.

Intervention

The most popular models to explain work-related stress are the effort-reward imbalance and the job demand-control model. According to the effort-reward imbalance model [2], work-related stress is generated by high effort (eg, pace of work and workload) and low reward received in return (eg, inadequate salary, promotion prospects, and job security). The job demand-control model [3] identifies high demand (eg, high workload) and low job decision latitude (eg, autonomy and control over the job) as factors that lead to high levels of job strain. This strain is known to be a risk factor for adverse health consequences, such as mental health problems [5], chronic pain [7], and cardiovascular disease [6]. Ideally, job strain should be reduced by changing adverse working conditions such as small rooms and bad equipment. As changing these may be difficult, stressors on an individual level such as inadequate coping strategies can also be addressed. Interventions based on Lazarus’s transactional model aim to empower the individual to reduce or modify problems at work (ie, high effort, low rewards, or low decision latitudes). This model identifies 2 strategies of coping with stressors: problem-oriented coping, to actively change or adapt stressors, and emotion-oriented coping, to cope with negative emotions due to stressors at the workplace. Thus, the iSMI is based on 2 main components: problem solving and emotion regulation. Problem solving is an evidence-based method for dealing with such problems and has been proven to be successful in improving mental health [28]. However, employees are frequently faced with unsolvable problems, which are associated with strong negative affective reactions and require effective regulation strategies. Improvement of emotion regulation skills has been shown to be both promising for reducing psychopathological symptoms [29] and a mechanism of change in previous studies using this iSMI [30]. Deficits in emotion regulation may also be an important factor for the development and persistence of mental health symptoms [31]. Yet, emotion-focused coping is regarded as the forgotten component, whereas problem-focused coping by means of problem-solving techniques is a well-established component of most cognitive-behavioral stress management trainings.

The iSMI is based on Lazarus's transactional model of stress and includes problem solving and emotion regulation. The intervention consists of 8 sessions composed of modules for psycho-education (session 1), problem solving (sessions 2 and 3), emotion regulation (sessions 4-6), planning for the future (session 7), and a booster session (session 8). In addition, participants could choose optional modules covering different topics, for example, time management, rumination and worrying, psychological detachment from work, and sleep hygiene. Each module takes approximately 45 to 60 min to complete. Participants were advised to complete 1 to 2 modules per week. Transfer tasks such as homework assignments were integrated into the intervention to help participants integrate learned skills into daily life. Participants received nontherapeutic feedback by an e-Coach after each completed module. E-Coaches had a degree in psychology, and feedbacks were based on a standardized manual on feedback writing. Participants could also opt in for an additional text message coach along the iSMI (eg, short relaxation exercises). A detailed description of the iSMI can be found elsewhere [32]. The clinical effectiveness of the iSMI has been positively evaluated in a series of RCTs [23,30,31,33,34].

Outcome Measures

Self-reported measures of stress and social functioning (PSS-10 and Short-Form Six-Dimension; SF-6D) were collected at baseline (T1), post treatment (T2; 7 weeks after randomization), and 6-month follow-up (T3) using a secured Web-based assessment system (AES, 256-bit encrypted).

Clinical Outcome

The level of perceived stress was measured by the PSS-10 [27]. Cronbach alphas indicated that the internal consistency ranged from .70 to .91 over different measurement points in this study [30]. Symptom-free status was operationalized as scoring 2 SDs below the PSS-10 sample mean at T1 (mean 25.52, SD 3.91) [23,35].

Quality-Adjusted Life Years

Quality-Adjusted Life Years (QALYs) were used as the primary outcome in the cost-utility analysis. QALYs were computed using the SF-6D [36]. A QALY gain of 0.5 indicates full health throughout the 6-month trial period. The SF-6D is more sensitive to change in mild conditions than the more commonly used EQ-5D and was used for the main analysis [37].

Resource Use and Costing

We assessed direct and indirect costs which occurred over the previous 3 months at baseline, and at 6-month follow-up. All costs were calculated in Euros for the reference year 2013 (index factor 1.04 based on the year 2010), referring to the German consumer price index [38]. Costs were converted to US dollar using the purchasing power parities reported by the Organization for Economic Cooperation and Development. For the reference year 2013, €1 was equated to US $1.29.

The Trimbos Institute and Institute of Medical Technology Questionnaire for Costs Associated with Psychiatric Illness (TiC-P) adapted to the German health care system was used [39]. This is a widely used and reliable instrument for collecting self-reported data on health care utilization and productivity losses in patients with mild to moderate mental health conditions [40-46]. The German version has been used in a number of health economic evaluations alongside randomized trials [21,41,42,44]. The standard unit cost prices were multiplied by the units of resource use for each participant. Multimedia Appendix 1 presents direct medical and direct nonmedical costs by health service type. Cumulated costs of the trial were estimated using the area under curve method to linearly interpolate 3 months costs as measured at each measurement point to cover the full follow-up period of 6 months [47].

Health Care Costs

Health care costs were calculated according to the guidelines of Kraut and Bock et al [48,49]. We included unit costs for a physician; a medical specialist; psychological services such as a psychiatrist and psychotherapist; and allied health services such as physiotherapy, massage, occupational therapy, as well as inpatient care and rehabilitation.

Medication

Unit costs of prescription drugs were calculated using the German register for pharmaceutical drugs *Rote Liste* [50]. The basis for calculating costs of prescribed medication is the pharmacy retail price accounting for a specific pharmacy and manufacture's discount. The discount rates vary between private and statutory health insurances [48]. Therefore, we weighted the mean costs of the 3 largest packages with the same agent based on the daily defined dose by the statutory population share (88,80% of the German population are statutorily insured).

Intervention Costs

The provider (GET.ON Institute GmbH) of the iSMI intervention GET.ON Stress estimated the current market price of the intervention at €299 (US $386) per participant. This flat tariff covers all costs for developing and hosting the intervention plus coaching of the participants. In general, it was assumed that every participant owned a computer, had access to the internet, and used the iSMI in their leisure time after working hours. Hence, these costs were not included.

Patient and Family Costs

Participants self-reported the cost of their out-of-pocket expenses (eg, for over-the-counter drugs). Direct nonmedical travel costs were calculated based on self-reported data that included the used method of transportation (ie, bus, taxi, or car) and round-trip distance to reach health care services. Each kilometer by car was valued at €0.30 [51]. Time spent by participants completing the intervention and/or receiving or waiting for treatment by a physician was considered part of their leisure time. The opportunity cost of leisure time, defined as the cost associated with the next best alternative use of a particular resource, was valued at €23.10 per hour. This was based on Bock et al's recommendations [48], which estimated these costs based on the average net wage of German employees plus their average pension and unemployment insurance contributions.

Costs incurred from a domestic help (help with daily chores) or production losses resulting from unpaid work such as informal care by friends and family were calculated using the substitution method. These costs were based on the average gross hourly wage earned by a domestic worker, as suggested by Bock et al [48]. This time was valued at €18.33 per hour.

Costs of Productivity Losses

Absenteeism costs were calculated by applying the human capital approach [52]. In doing so, the number of work loss days was multiplied by the participant's average gross daily wage based on their reported monthly salary. In addition, participants reported the number of workdays for which they reported lesser efficiency. On the basis of the Osterhaus method [53], these days were multiplied by an inefficiency score, which resulted in lost-workday equivalents due to presenteeism. Subsequently, based on self-reported monthly salary, their gross wages per day were calculated and used to calculate the costs that occurred due to presenteeism.

Statistical Analysis

This study was powered to detect a mean difference of d=0.35 in the primary outcome (PSS) between the groups at post measurement. Cost data are usually heavily skewed to the right, with large variance requiring very large sample sizes to test the statistical significance of cost differences. Instead, we adopted a probabilistic decision-making approach for our economic analyses [54]. This procedure takes the stochastic uncertainty of the trial data into account [55] and informs the decision makers on probabilities rather than statistical significance. Due to the 6-month follow-up period, no discounting was applied.

All analyses were conducted in accordance with the intention-to-treat (ITT) principle. Missing clinical outcome data were imputed using a Markov Chain Monte Carlo multivariate imputation algorithm with 10 estimations per missing value.

Missing cost data were imputed using the regression imputation procedure implemented in Stata to obtain the required predicted values. Predictors of outcome and dropout were identified via (logistic) regression. Differences in PSS score and symptom-free status between groups were assessed at follow-up using the Chi-square test. At baseline, mean SF-6D utility values were similar in both groups (WLC: mean 0.65, SD 0.08 and iSMI: mean 0.65, SD 0.11). Therefore, no baseline adjustments were made when calculating QALYs. Differences in QALYs between iSMI and WLC were assessed using independent samples t tests.

Analysis of Cost-Effectiveness and Cost-Utility

For the cost-effectiveness analyses, the incremental cost-effectiveness ratio (ICER) was calculated as incremental costs per unit of effect (QALY and symptom-free status). Symptom-free status is meaningful for decision makers and was used as the preferred effect measure as there was no difference between beta coefficients from an OLS regression on the binary outcome compared with beta coefficients from a linear probability model in GLM (GLM: beta=.36, P<.001 and OLS: beta=.36, P<.001).

The ICER was calculated as ICER=$(costs_{\text{iSMI}}-costs_{\text{WLC}})/(effects_{\text{iSMI}}-effects_{\text{WLC}})$, where costs are the cumulated *costs* over the 6-month period and *effect* are QALY gains or symptom-free status.

Stochastic uncertainty in the ICER was handled using nonparametric bootstrapping, which is a resampling technique applied to the trial data, which generates 5000 simulations of the ICER. The incremental costs and incremental effects were obtained under a bootstrapped seemingly unrelated regression equations model and allowed for correlated residuals of the cost and effect equations [56]. The 5000 bootstrap replications of costs and effects were also used to obtain 95% CIs based on the percentile method.

In a next step, the simulated ICERs were plotted in a cost-effectiveness plane. On the plane, incremental effects are depicted on the horizontal x-axis and the incremental costs on the vertical y-axis. Each dot in the cost-effectiveness plane represents 1 bootstrapped ICER.

The willingness-to-pay (WTP) threshold reflects the maximum amount the society would be willing to pay for a health benefit (eg, a symptom-free person or a QALY gained). As the WTP ceiling for gaining 1 unit of health (eg, gaining 1 QALY or obtaining symptomatic remission in 1 person) is an unknown quantity, a cost-effectiveness acceptability curve was presented, which displays the probability of the intervention being cost-effective for 1 additional unit of health gained at varying WTP ceilings. All analyses were performed using Stata version 13 [57].

Sensitivity Analyses

The robustness of the outcomes was assessed using several sensitivity analyses. First, we used the EQ-5D-3L (European Quality of Life 5 Dimensions 3 Level) instrument [58] for the calculation of QALYs. Second, there is uncertainty regarding the cost of the intervention due to changing demand. Therefore, we conducted sensitivity analyses assuming higher and lower interventions costs (±€100). Third, inpatient costs tend to be very high, but they were only reported by a few participants (n=9, 3.4%). Such outliers may lead to distorted outcomes results, so they were removed in the final sensitivity analysis.

Results

Sample

Multimedia Appendix 2 presents the baseline characteristics. Interested participants were recruited from the general working population via mass media (eg, newspaper articles and television) and with the aid of a health insurance company (BARMER) within their occupational health management program. An open-access website [59] was used to sign-up for study participation. The sample predominately consisted of full-time employed middle-aged women living with a partner. A comprehensive description of the study sample and the participant flow can be found elsewhere [23]. We did not observe any clinically relevant baseline differences between study conditions.

Study Dropouts

The study attrition was low: 10.6% (28/264) of participants did not complete the 6-month follow-up assessment. The dropout rates between the groups, with 12.8% (17/132) in the iSMI condition and 8.33% (11/132) in the WLC condition, did not differ significantly (χ^2_1=1.4 *P*=.23).

Outcome Measures

The iSMI improved by 9.75 (SD 6) PSS-10 stress units between pre and 6-month follow-up, whereas the WLC improved by 3.0 units (SD 6) PSS. Differences regarding symptom-free status based on the PSS-10 between groups were assessed at follow-up (iSMI: 79/132, 59.8%; WLC: 31/132, 23.5%; χ^2_1=35.9; *P*<.001; NNT (Number needed to treat) =2.75, 95% CI 2.11-3.95) [23]. However, the intervention and the WLC did not differ significantly in terms of SF-6D QALY gains (iSMI=0.35, SD 0.04 vs WLC=0.35, SD 0.35; t_{262}=−1.625; *P*=.10).

Costs

At baseline, mean total costs were €3239 (US $4178) in the iSMI and €3183 (US $4178) in the WLC, which is only a small difference of €56 (US $72), indicating that randomization had resulted in a well-balanced trial. Table 1 presents the average 6-month accumulated per-participant costs by study condition. The costs are clustered into health care costs, patient and family costs, and costs stemming from productivity losses. After 6 months, total incremental costs were €380 (US $490); thus, the iSMI group had less costs than WLC (iSMI: €5258 and WLC: €5642). Health care costs were, on average, higher in the iSMI group compared with WLC. Hospital admissions were a major cost driver. Regarding the patient and family costs, the iSMI had less costs than WLC. Informal care was decreased by €241 for the iSMI. Finally, productivity losses produced the highest cost differences of €487, exceeding the intervention costs, meaning that the iSMI produced less cost than WLC.

Cost-Effectiveness

Table 2 shows the incremental costs, effects, and cost-effectiveness ratios based on 5000 bootstrapped simulations. The bootstrapped ICER for symptom-free status on the PSS-10 was dominant. The cost-effectiveness plane is shown in Figure 1. The majority (70%) of the bootstrapped ICERs fell in the south-east quadrant, indicating a 70% probability that the intervention produces greater health at lower costs than WLC. Hence, the iSMI intervention dominates the WLC condition from a societal perspective. The remaining 30% of ICERs fell in the north-east quadrant, indicating a 30% probability that the intervention produces greater health at greater costs than WLC. Figure 2 presents the cost-effectiveness acceptability curve. If the decision maker is willing to pay €1000 and €3000 for gaining a symptom-free person, the intervention's probability of being more cost-effective than WLC rises to 85% and 97%, respectively.

Table 1. Average costs per participant (in €) by condition at 6-months follow-up (area under the curve, intention-to-treat-sample, N=264).

Cost category	Internet-based stress management intervention (n=132), mean (SD)	Waitlist control condition (n=132), mean (SD)	Incremental costs, difference
Health care costs (€)			
Intervention	299 (Reflects a fixed price)	0 (Reflects a fixed price)	299
Physician services	132 (139)	147 (175)	−15
Psychological services	111 (291)	209 (468)	−98
Hospital in-patient	342 (2222)	188 (1237)	154
Hospital semiresidential	234 (1444)	77 (798)	157
Rehabilitation	8 (41)	89 (658)	−81
Nonphysician services	167 (293)	174 (314)	−7
Prescription drugs	50 (97)	56 (105)	−6
Patient and family costs (€)			
Over the counter drugs	48 (88)	48 (78)	0
Opportunity costs	485 (754)	526 (892)	−42
Travel expenses	27 (48)	49 (94)	−21
Domestic help or informal care	424 (1213)	665 (1327)	−241
Productivity losses (€)			
Absenteeism	1346 (2184)	1655 (3436)	−309
Presenteeism	1578 (1471)	1756 (1849)	−178
Total costs (€)[a]	5258 (5493)	5642 (6000)	−384

[a] Due to rounding, numbers presented may not add up precisely to the totals provided.

Table 2. Results of the main and sensitivity analysis based on 5000 bootstrap simulations. Costs are expressed in 2013 Euros.

Analysis and outcome	Incremental costs, € (95% CI)	Incremental effects, points (95% CI)	Incremental cost-effectiveness ratio, €/points (95% CI)[a]	Distribution over the cost-effectiveness plane, %			
				North-east quadrant[b]	South-east quadrant[c]	South-west quadrant[d]	North-west quadrant[e]
Main analysis							
Perceived stress (range 0-40)	−386 (−1794 to 1006)	6.27 (4.9 to 7.7)[f]	Dominant (dominant to 171)	30	70	__[g]	—
Symptom-free status (0/1)	−386 (−1794 to 1006)	0.362 (0.25 to 0.47)[f]	Dominant (dominant to 3360)	30	70	—	—
QALYs[h] (range: 0-1)	−386 (−1794 to 1006)	0.0074 (−0015 to 0.016)	Dominant[i]	26	69	2	3
Sensitivity analysis 1[j]							
Perceived stress (range 0-40)	−616 (−1731 to 485)	6.27 (4.9 to 7.7)[f]	Dominant (dominant to 81)	13	87	—	—
Symptom-free status (0/1)	−616 (−1731 to 485)	0.362 (0.25 to 0.47)[f]	Dominant (dominant to 1415)	13	87	—	—
QALYs (range: 0-1)	−616 (−1731 to 485)	0.0074 (−0015 to 0.016)	Dominant[i]	12	83	2	3
Sensitivity analysis 2[k] , €+100 added to intervention costs							
Perceived stress (range 0-40)	−286 (−1694 to 1106)	6.27 (4.9 to 7.7)[f]	Dominant (dominant to 187)	34	66	—	—
Symptom-free status (0/1)	−286 (−1694 to 1106)	0.362 (0.25 to 0.47)[f]	Dominant (dominant to 3419)	34	66	—	—
QALYs (range: 0-1)	−286 (−1694 to 1106)	0.0075 (−0015 to 0.016)	Dominant[i]	31	64	2	3
Sensitivity analysis 2[k] , €-100 added to intervention costs							
Perceived stress (range 0-40)	−486 (−1894 to 906)	6.27 (4.9 to 7.7)[f]	Dominant (dominant to 155)	24	76	—	—
Symptom-free status (0/1)	−486 (−1894 to 906)	0.362 (0.25 to 0.47)[f]	Dominant (dominant to 2764)	24	76	—	—
QALYs (range: 0-1)	−486 (−1894 to 906)	0.0075 (−0015 to 0.016)	Dominant[i]	22	73	2	3
Sensitivity analysis 3[l]							
QALYs (range: 0-1)	−386 (−1794 to 1006)	0.00186 (−0.010 to 0.014)	Dominant[i]	49	14	22	16

[a] In line with the best practice ISPOR guidelines on ‘Model Parameter Estimation and Uncertainty’ we did not report negative incremental cost-effectiveness ratios (ICERs) as they are meaningless. Instead we used the term dominant which implies that the intervention has a higher effect and less cost compared with the WLC.

[b] The north-east quadrant of the CE plane, indicating that intervention is more effective and more costly.

[c] The south-east quadrant of the CE plane, indicating that intervention is more effective and less costly.

[d] The south-west quadrant of the CE plane, indicating that intervention is less effective and less costly.

[e] The north-west quadrant of the CE plane, indicating that intervention is less effective and more costly.

[f] *P*<.05.

[g] The distribution of the ICERs (N=5000) sums to 100%. If the distribution only falls into 2 quadrants, there will not be any ICER in the other 2 quadrants (= 0%).

[h] QALYs: quality-adjusted life years.

[i] A dependably accurate 95% confidence interval for this distribution cannot be defined because there is no line through the origin that excludes alpha/2 of the distribution.

[j] Sensitivity analysis 1 analyses not including in-patient care.

[k] Sensitivity analysis 2 analyses adding €±100 of intervention costs.

[l] Sensitivity analysis 3 analyses for EQ5D quality-adjusted life years.

Figure 1. Scatterplot of 5000 replicates of the incremental cost-effectiveness ratio (mean differences in costs and symptom-free status) on the cost-effectiveness plane: internet-based stress-management intervention versus waitlist control condition.

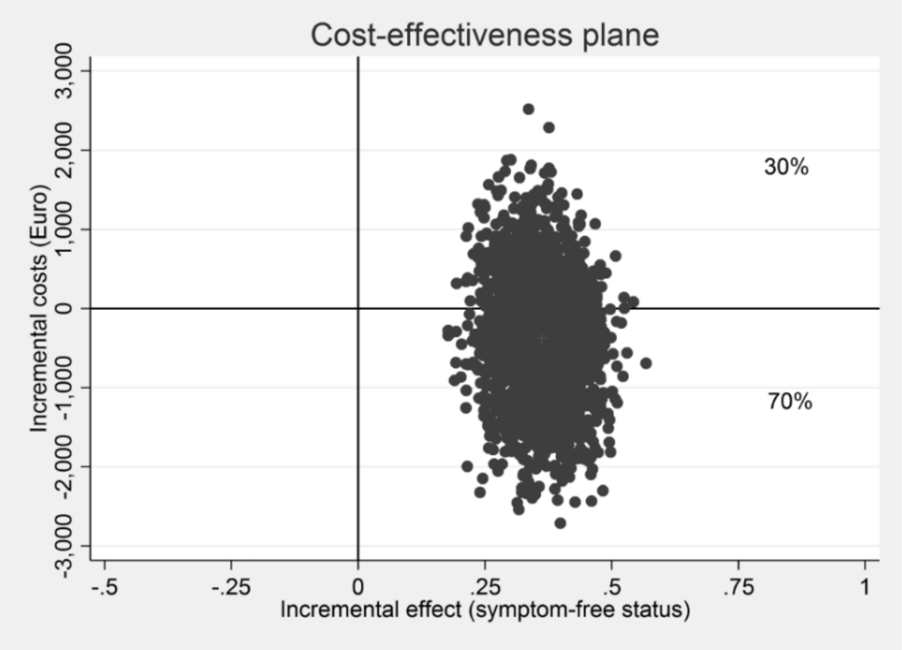


Figure 2. Cost-effectiveness acceptability curve showing the probability of the internet-based stress-management intervention being cost-effective at varying willingness-to-pay ceilings (based on 5000 replicates of the incremental cost-effectiveness ratio using mean differences in costs and symptom-free status).

Cost-Utility

The ICER based on QALY gains showed a small health benefit (approximately 0.001 QALYs gained) for lower mean costs (€386; US $498). Of the simulated ICERs, 69% (as seen in Figure 3) fell in the south-east quadrant, reflecting the intervention's probability of dominating WLC, whereas 26% fell in the north-east quadrant , indicating higher costs and health gains, and 2% fell in the south-west quadrant and 3% in north-west quadrant. Assuming a WTP of €10,000 and €20,000 for gaining 1 QALY, the probability rose to 73% and 76%, respectively (Figure 4).

Figure 3. Scatterplot of 5000 replicates of the incremental cost-effectiveness ratio (mean differences in costs and quality-adjusted life years based on the Short-Form Six-Dimension) on the cost-effectiveness plane: internet-based stress-management intervention versus waitlist control condition. QALY: quality-adjusted life years; SF-6D: Short-Form Six-Dimension.

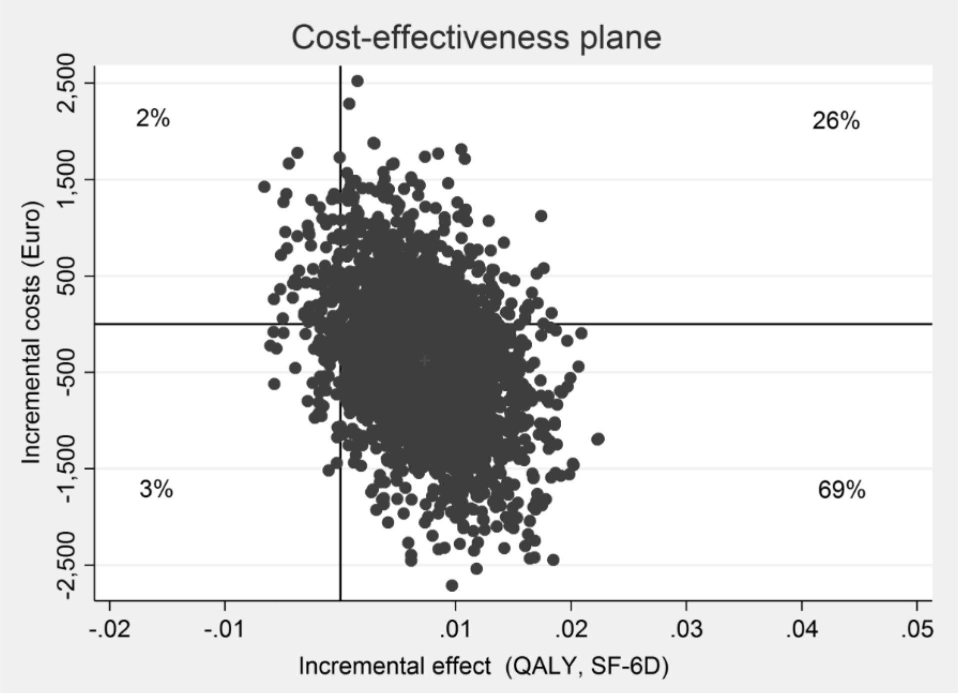
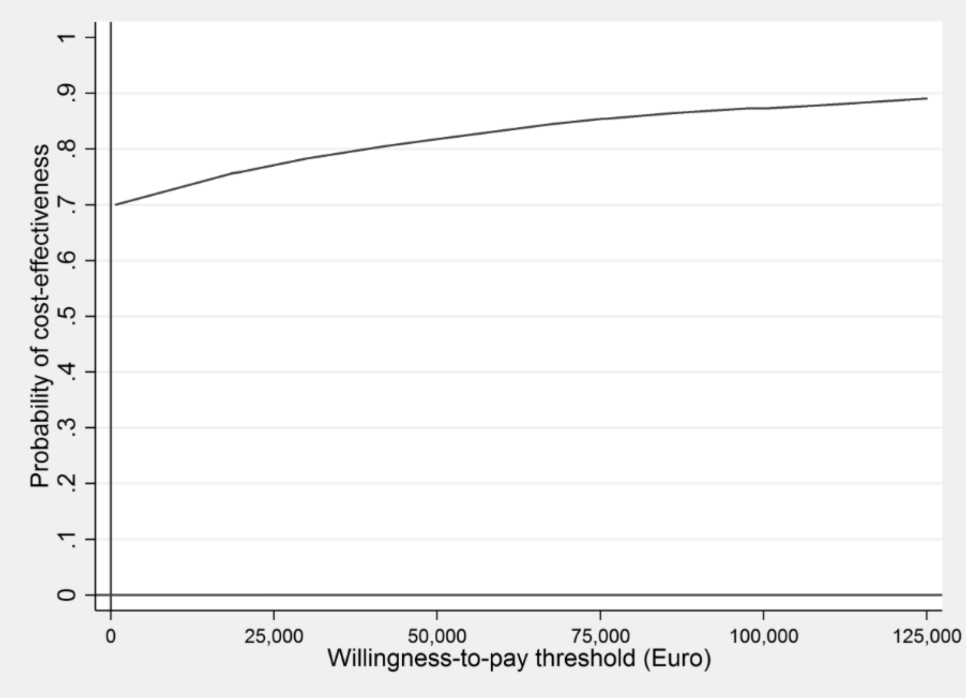


Figure 4. Cost-effectiveness acceptability curve showing the probability of the internet-based stress-management intervention being cost-effective at varying willingness-to-pay ceilings (based on 5000 replicates of the incremental cost-effectiveness ratio using mean differences in costs and quality-adjusted life years based on the Short-Form Six-Dimension). QALY: quality-adjusted life years; SF-6D: Short-Form Six-Dimension.



Sensitivity Analyses

Using the EQ-5D-3L resulted in a smaller incremental QALY gain in favor of the intervention group (0.28 QALY, SD 0.05) compared with WLC (0.28 QALY, SD 0.05), which was not statistically significant (t_{262}=−0.296; *P*=.77). This is in line with available evidence that the EQ-5D-3L suffers from ceiling effects in milder conditions [37]. Nevertheless, at a WTP of €20,000 for gaining a QALY, the probability of being cost-effective was 71%.

As inpatient costs were reported from only a few participants but were associated with high costs, these costs might have distorted the results. Excluding these costs led to higher ICERs for both outcomes (eg, symptom-free status and QALYs). The probability of being cost-effective rose to 86% and 96% at a

WTP of €0 and €1000 with regard to symptom-free status, and 86% and 90% for gaining a QALY, respectively.

Increasing and subsequently reducing the intervention costs by €100 led to a 66% and 76% probability that the intervention produces a greater health gain at lower costs than WLC with regard to symptom-free status and 1-point improvement.

Discussion

Principal Findings

This study evaluated the cost-effectiveness and cost-utility of a Web-based guided self-help intervention for employees with elevated stress levels aimed at reducing perceived stress compared with WLC from the societal perspective. The intervention had a significant and favorable effect on perceived stress after 6 months and a high probability of being cost-effective compared with the control condition. The overall conclusion of this study does not change when using any of the assumptions, as explored in the sensitivity analyses.

Strengths and Limitations

First, we had missing data, which were handled using imputation techniques to perform an ITT analysis of both effects and costs [60]. As dropout rate was very low (12.8% for the iSMI and 8.33% for the WLC at 6 months), it is unlikely that this has biased the results substantially. Second, the costs and effects were only evaluated over a 6 months period. Hence, we cannot draw any conclusions about long-term effects. Third, self-reported costs and effects might have led to social desirability and/or recall bias. Nonetheless it seems unlikely that this bias differed systematically between groups due to absent baseline differences. Fourth, approaches used for cost estimation of lost productivity are based on the participants' wages which do not reflect the average wages in the general population. Fifth, a waitlist control group design with unrestricted access to treatment as usual was chosen, which causes participants to be less motivated to initiate health-related behavior changes and thus over-accentuates effects [61]. Sixth, the majority of the sample was female, which is a common feature of mental health internet-based interventions [62]. The gender imbalance might limit the generalizability of study findings. Finally, the use of behavioral interventions does not result in improved working conditions that could cause less job strain. However, the potential of workplace-related interventions is often not fully utilized, and hence, such interventions are not systematically implemented. Thus, we recommend a combined implementation to design healthy working conditions.

Comparison With Findings From Other Studies

The results of this study with an effect size of *d*=0.83 [23] on perceived stress are in line with the meta-analytic evidence (pooled effect size of *d*=0.43, 95% CI: 0.31-0.54) [16].

In addition, some evidence exists for the economic benefits of stress management and internet-based interventions to reduce depressive symptoms in employees. However, to the best of our knowledge, this study is the first study to evaluate the cost-effectiveness of a Web-based guided self-help intervention for employees with elevated stress levels.

Jacobsen et al evaluated the costs of a self- and professional-administered stress-management intervention not delivered over the internet in patients undergoing chemotherapy compared with usual care [63]. Lower costs and statistically higher quality of life outcomes were found in the intervention group. Hedman et al compared behavioral stress management with iCBT for treatment of severe health anxiety. The iSMI resulted in lower costs but was not considered cost-effective [22].

In a Web-based intervention by Geraedts [64], the probabilities of cost-effectiveness were 0.62 (societal perspective) and 0.55 (employer's perspective) compared with WLC in employees with depressive symptoms. The intervention was not judged cost-effective. Besides that, the reduction of depressive symptoms was rather small (*d*=0.16) [65] compared with our study (*d*=0.64) [23] at post measurement. However, Buntrock et al reported an effect size of *d*=0.69 for a Web-based intervention for the prevention of depression. This intervention has an acceptable likelihood of being more cost-effective than enhanced usual care [21]. Focusing on perceived stress rather than on depressive symptoms in employees seems to be a cost-effective strategy to reduce the mental burden.

Clinical Implications

The results of this study support the idea that iSMIs could be a promising cost-effective strategy in reducing adverse effects of persistent stress in the workplace. Intervention costs were mainly driven by psychologists who acted as e-Coaches. Yet, studies showed that iSMIs are also effective when delivered in a less costly adherence-focused guidance and pure self-help format [33]. However, meta-analytic evidence shows that guidance yields higher effect sizes [66]. Therefore, the cost-effectiveness of guided versus unguided iSMI needs to be evaluated.

Long-term costs caused by persistent stress, such as staff turnover or mental health disorder onsets, were not taken into account. Future studies should investigate the long-term economic effects of iSMIs. The sample consisted predominately of middle-aged women. Future research should focus on the general German working population regarding recruitment, implementation, and dissemination.

Conclusions

This study demonstrated that this iSMI has a high probability of being cost-effective in reducing stress levels when compared with WLC. Given the increasing stress in the workplace and the small number of people who are reached via available health care services [67], it would be worthwhile to integrate such iSMIs into routine occupational health care, which conventionally only consists of face-to-face therapy by occupational health physicians.

Acknowledgments

The authors gratefully acknowledge the contribution of all participants, research assistants, therapists, and all others who contributed to this study. Funding was received from the European Union (project number: EFRE: CCI 2007DE161PR001) and the BARMER (German statutory health insurance company). The funders did not have a role in study design, data collection, analyses or interpretation of the results, or the decision to publicize the study results.

Authors' Contributions

MB obtained the funding for this trial and DDE obtained the funding for the analysis. EH was responsible for the study administration. FK analyzed the data and drafted the manuscript on the health-economic evaluation, and DDE supervised this process. All authors contributed to the further writing of the manuscript and approved the final version of the manuscript.

Conflicts of Interest

DDE, DL, EH, and MB are stakeholders of the "Institute for Online Health Trainings," which aims to transfer scientific knowledge related to this research into routine health care.

Multimedia Appendix 1

Unit costs for the type of health service utilized by the participants.

[PDF File (Adobe PDF File), 14KB - mental_v6i4e10866_app1.pdf]

Multimedia Appendix 2

Demographic characteristics: means/counts, standard deviations/percentages at baseline.

[PDF File (Adobe PDF File), 14KB - mental_v6i4e10866_app2.pdf]

References

1. Eurofound. Developments in Working Life in EuropeurWORK Annual Review. Luxembourg: Publications Office of the European Union; 2016.
2. Siegrist J. Adverse health effects of high-effort/low-reward conditions. J Occup Health Psychol 1996 Jan;1(1):27-41. [doi: 10.1038/372669a0] [Medline: 9547031]
3. Karasek Jr RA. Job demands, job decision latitude, and mental strain: implications for job redesign. Adm Sci Q 1979 Jun;24(2):285-308. [doi: 10.2307/2392498]
4. Åkerstedt T, Orsini N, Petersen H, Axelsson J, Lekander M, Kecklund G. Predicting sleep quality from stress and prior sleep--a study of day-to-day covariation across six weeks. Sleep Med 2012 Jun;13(6):674-679. [doi: 10.1016/j.sleep.2011.12.013] [Medline: 22621983]
5. Theorell T, Hammarström A, Aronsson G, Träskman BL, Grape T, Hogstedt C, et al. A systematic review including meta-analysis of work environment and depressive symptoms. BMC Public Health 2015 Aug 1;15:738 [FREE Full text] [doi: 10.1186/s12889-015-1954-4] [Medline: 26232123]
6. Dragano N, Siegrist J, Nyberg ST, Lunau T, Fransson EI, Alfredsson L, IPD-Work Consortium. Effort–reward imbalance at work and incident coronary heart disease: a multicohort study of 90,164 individuals. Epidemiology 2017 Dec;28(4):619-626 [FREE Full text] [doi: 10.1097/EDE.0000000000000666] [Medline: 28570388]
7. Sihawong R, Sitthipornvorakul E, Paksaichol A, Janwantanakul P. Predictors for chronic neck and low back pain in office workers: a 1-year prospective cohort study. J Occup Health 2016;58(1):16-24 [FREE Full text] [doi: 10.1539/joh.15-0168-OA] [Medline: 26498979]
8. Niedhammer I, Chastang J, Sultan-Taïeb H, Vermeylen G, Parent-Thirion A. Psychosocial work factors and sickness absence in 31 countries in Europe. Eur J Public Health 2013 Aug;23(4):622-629. [doi: 10.1093/eurpub/cks124] [Medline: 23002241]
9. Uegaki K, Bakker I, de Bruijne M, van der Beek A, Terluin B, van Marwijk H, et al. Cost-effectiveness of a minimal intervention for stress-related sick leave in general practice: results of an economic evaluation alongside a pragmatic randomised control trial. J Affect Disord 2010 Jan;120(1-3):177-187. [doi: 10.1016/j.jad.2009.04.012] [Medline: 19439362]
10. Rebergen DS, Bruinvels DJ, van Tulder MW, van der Beek AJ, van Mechelen W. Cost-effectiveness of guideline-based care for workers with mental health problems. J Occup Environ Med 2009 Mar;51(3):313-322. [doi: 10.1097/JOM.0b013e3181990d8e] [Medline: 19225416]
11. Kalia M. Assessing the economic impact of stress--the modern day hidden epidemic. Metabolism 2002 Jun;51(6 Suppl 1):49-53. [doi: 10.1053/meta.2002.33193] [Medline: 12040542]
12. Hassard J, Teoh KR, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: a systematic review. J Occup Health Psychol 2018 Jan;23(1):1-17. [doi: 10.1037/ocp0000069] [Medline: 28358567]

13. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. J Occup Health Psychol 2008 Jan;13(1):69-93. [doi: 10.1037/1076-8998.13.1.69] [Medline: 18211170]
14. Hilton MF, Whiteford HA, Sheridan JS, Cleary CM, Chant DC, Wang PS, et al. The prevalence of psychological distress in employees and associated occupational risk factors. J Occup Environ Med 2008 Jul;50(7):746-757. [doi: 10.1097/JOM.0b013e31817e9171] [Medline: 18617830]
15. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
16. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web-and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
17. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
18. Blankers M, Nabitz U, Smit F, Koeter MW, Schippers GM. Economic evaluation of internet-based interventions for harmful alcohol use alongside a pragmatic randomized controlled trial. J Med Internet Res 2012;14(5):e134 [FREE Full text] [doi: 10.2196/jmir.2052] [Medline: 23103771]
19. Graham AL, Chang Y, Fang Y, Cobb NK, Tinkelman DS, Niaura RS, et al. Cost-effectiveness of internet and telephone treatment for smoking cessation: an economic evaluation of The iQUITT Study. Tob Control 2013 Nov;22(6):e11 [FREE Full text] [doi: 10.1136/tobaccocontrol-2012-050465] [Medline: 23010696]
20. Nordgren LB, Hedman E, Etienne J, Bodin J, Kadowaki A, Eriksson S, et al. Effectiveness and cost-effectiveness of individually tailored internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: a randomized controlled trial. Behav Res Ther 2014 Aug;59:1-11 [FREE Full text] [doi: 10.1016/j.brat.2014.05.007] [Medline: 24933451]
21. Buntrock C, Berking M, Smit F, Lehr D, Nobis S, Riper H, et al. Preventing depression in adults with subthreshold depression: health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. J Med Internet Res 2017 Jan 4;19(1):e5 [FREE Full text] [doi: 10.2196/jmir.6587] [Medline: 28052841]
22. Hedman E, Andersson E, Ljótsson B, Axelsson E, Lekander M. Cost effectiveness of internet-based cognitive behaviour therapy and behavioural stress management for severe health anxiety. BMJ Open 2016;6(4):e009327 [FREE Full text] [doi: 10.1136/bmjopen-2015-009327] [Medline: 27113231]
23. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. J Med Internet Res 2016;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
24. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, CHEERS Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Br Med J 2013 Mar 25;346:f1049 [FREE Full text] [doi: 10.1136/bmj.f1049] [Medline: 23529982]
25. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. Value Health 2005;8(5):521-533 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.00045.x] [Medline: 16176491]
26. Datinf Randlist - Software to create Randomisation Lists. URL: http://randomisation.eu/ [accessed 2019-04-12] [WebCite Cache ID 77ZhbpKfa]
27. Lesage F, Berjot S, Deschamps F. Psychometric properties of the French versions of the perceived stress scale. Int J Occup Med Environ Health 2012 Jun;25(2):178-184 [FREE Full text] [doi: 10.2478/S13382-012-0024-8] [Medline: 22528542]
28. Malouff JM, Thorsteinsson EB, Schutte NS. The efficacy of problem solving therapy in reducing mental and physical health problems: a meta-analysis. Clin Psychol Rev 2007 Jan;27(1):46-57. [doi: 10.1016/j.cpr.2005.12.005] [Medline: 16480801]
29. Berking M, Ebert D, Cuijpers P, Hofmann SG. Emotion regulation skills training enhances the efficacy of inpatient cognitive behavioral therapy for major depressive disorder: a randomized controlled trial. Psychother Psychosom 2013;82(4):234-245. [doi: 10.1159/000348448] [Medline: 23712210]
30. Ebert D, Lehr D, Heber E, Riper H, Cuijpers P, Berking M. Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand J Work Environ Health 2016 Dec 1;42(5):382-394 [FREE Full text] [doi: 10.5271/sjweh.3573] [Medline: 27249161]
31. Ebert DD, Lehr D, Smit F, Zarski A, Riper H, Heber E, et al. Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial. BMC Public Health 2014;14:807 [FREE Full text] [doi: 10.1186/1471-2458-14-807] [Medline: 25099533]
32. Heber E, Ebert DD, Lehr D, Nobis S, Berking M, Riper H. Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial. BMC Public Health 2013;13:655 [FREE Full text] [doi: 10.1186/1471-2458-13-655] [Medline: 23855376]

33. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
34. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
35. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1016/S0005-7894(86)80060-0] [Medline: 2002127]
36. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
37. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004 Sep;13(9):873-884. [doi: 10.1002/hec.866] [Medline: 15362179]
38. Federal Statistical Office of Germany. 2015. Consumer Price Indices URL: https://www.destatis.de/EN/FactsFigures/NationalEconomyEnvironment/Prices/ConsumerPriceIndices/ConsumerPriceIndices.html, [accessed 2016-08-30] [WebCite Cache ID 6k9PhqEof]
39. Van Hakkaart RL, Van Straten A, Donker M, Tiemens B. Manual Trimbos/iMTA Questionnaire for Costs Associated with Psychiatric Illness (TIC-P). Rotterdam: Institute for Medical Technology Assessment; Jan 2002.
40. Bouwmans C, de Jong K, Timman R, Zijlstra-Vlasveld M, van der Feltz-Cornelis C, Tan SS, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res 2013;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]
41. Nobis S, Lehr D, Ebert DD, Berking M, Heber E, Baumeister H, et al. Efficacy and cost-effectiveness of a web-based intervention with mobile phone support to treat depressive symptoms in adults with diabetes mellitus type 1 and type 2: design of a randomised controlled trial. BMC Psychiatry 2013;13:306 [FREE Full text] [doi: 10.1186/1471-244X-13-306] [Medline: 24238346]
42. Lin J, Lüking M, Ebert D, Buhrman M, Andersson G, Baumeister H. Effectiveness and cost-effectiveness of a guided and unguided internet-based Acceptance and Commitment Therapy for chronic pain: study protocol for a three-armed randomised controlled trial. Internet Interv 2015 Mar;2(1):7-16 [FREE Full text] [doi: 10.1016/j.invent.2014.11.005]
43. Buntrock C, Ebert DD, Lehr D, Cuijpers P, Riper H, Smit F, et al. Evaluating the efficacy and cost-effectiveness of web-based indicated prevention of major depression: design of a randomised controlled trial. BMC Psychiatry 2014;14:25 [FREE Full text] [doi: 10.1186/1471-244X-14-25] [Medline: 24485283]
44. Thiart H, Ebert DD, Lehr D, Nobis S, Buntrock C, Berking M, et al. Internet-based cognitive behavioral therapy for insomnia: a health economic evaluation. Sleep 2016 Oct 01;39(10):1769-1778 [FREE Full text] [doi: 10.5665/sleep.6152] [Medline: 27450686]
45. Boß L, Lehr D, Berking M, Riper H, Schaub MP, Ebert DD. Evaluating the (cost-)effectiveness of guided and unguided Internet-based self-help for problematic alcohol use in employees--a three arm randomized controlled trial. BMC Public Health 2015 Oct 12;15:1043 [FREE Full text] [doi: 10.1186/s12889-015-2375-0] [Medline: 26458872]
46. Sander L, Paganini S, Lin J, Schlicker S, Ebert DD, Buntrock C, et al. Effectiveness and cost-effectiveness of a guided internet- and mobile-based intervention for the indicated prevention of major depression in patients with chronic back pain-study protocol of the PROD-BP multicenter pragmatic RCT. BMC Psychiatry 2017 Jan 21;17(1):36 [FREE Full text] [doi: 10.1186/s12888-017-1193-6] [Medline: 28109247]
47. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. Br Med J 1990 Jan 27;300(6719):230-235 [FREE Full text] [doi: 10.1136/bmj.300.6719.230] [Medline: 2106931]
48. Bock J, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen 2015 Jan;77(1):53-61. [doi: 10.1055/s-0034-1374621] [Medline: 25025287]
49. Krauth C. [Methods of cost determination in the health economic evaluation]. Gesundh ökon Qual manag 2010 Oct 27;15(05):251-259. [doi: 10.1055/s-0029-1245680]
50. Rote Liste. [Red List® 2013 Book Edition - Single Issue: Drug List for Germany (Including EU Approvals and Certain Medical Devices)]. Frankfurt/Main: Rote Liste Service GmbH; 2013.
51. [Income Tax Code]. Berlin: Bundesministerium der Justiz (Federal Office of Justice); 2008.
52. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes, Fourth Edition. Oxford: Oxford University press; 2015.
53. Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare resource and lost labour costs of migraine headache in the US. Pharmacoeconomics 1992 Jul;2(1):67-76. [doi: 10.2165/00019053-199202010-00008] [Medline: 10146980]
54. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ 1994 Oct;3(5):309-319. [doi: 10.1002/hec.4730030505] [Medline: 7827647]
55. Briggs A, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technol Assess 1999;3(2):1-134 [FREE Full text] [Medline: 10448202]

56. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. Health Econ 2004 May;13(5):461-475. [doi: 10.1002/hec.843] [Medline: 15127426]
57. Stata: Release 13. Statistical Software.College Station. TX: StataCorp LP; 2013.
58. EuroQol Group. EuroQol—A new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
59. GesundheitsTraining.Online. URL: http://www.geton-training.de/ [accessed 2019-04-12] [WebCite Cache ID 77Zhlw7La]
60. Buchberger B, Heymann R, Huppertz H, Friepörtner K, Pomorin N, Wasem J. The effectiveness of interventions in workplace health promotion as to maintain the working capacity of health care personal. GMS Health Technol Assess 2011;7:Doc06 [FREE Full text] [doi: 10.3205/hta000097] [Medline: 22031811]
61. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. Acta Psychiatr Scand 2014 Sep;130(3):181-192. [doi: 10.1111/acps.12275] [Medline: 24697518]
62. Crisp DA, Griffiths KM. Participating in online mental health interventions: who is most likely to sign up and why? Depress Res Treat 2014;2014:790457 [FREE Full text] [doi: 10.1155/2014/790457] [Medline: 24804089]
63. Jacobsen PB, Meade CD, Stein KD, Chirikos TN, Small BJ, Ruckdeschel JC. Efficacy and costs of two forms of stress management training for cancer patients undergoing chemotherapy. J Clin Oncol 2002 Jun 15;20(12):2851-2862. [doi: 10.1200/JCO.2002.08.301] [Medline: 12065562]
64. Geraedts AS, van Dongen JM, Kleiboer AM, Wiezer NM, van Mechelen W, Cuijpers P, et al. Economic evaluation of a web-based guided self-help intervention for employees with depressive symptoms: results of a randomized controlled trial. J Occup Environ Med 2015 Jun;57(6):666-675. [doi: 10.1097/JOM.0000000000000423] [Medline: 25741795]
65. Geraedts AS, Kleiboer AM, Wiezer NM, van Mechelen W, Cuijpers P. Short-term effects of a web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014 May;16(5):e121 [FREE Full text] [doi: 10.2196/jmir.3185] [Medline: 24800966]
66. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
67. Levi L. Bridging the science–policy and policy–implementation gaps. In: Cooper CL, Campbell Quick J, editors. The Handbook of Stress and Health: A Guide to Research and Practice. Chichester: John Wiley & Sons; Apr 19, 2017:5-23.

Abbreviations

iCBT: internet-based cognitive behavior therapy
ICER: incremental cost-effectiveness ratio
iSMI: internet-based stress management interventions
ITT: intention-to-treat
PSS-10: Perceived Stress Scale
QALYs: quality-adjusted life years
RCT: randomized controlled trial
WLC: waitlist control condition
WTP: willingness-to-pay

Edited by G Eysenbach; submitted 04.05.18; peer-reviewed by T Donker, Y Lee, P Jimenez; comments to author 30.08.18; revised version received 02.11.18; accepted 30.12.18; published 15.05.19.

Please cite as:
Kählke F, Buntrock C, Smit F, Berking M, Lehr D, Heber E, Funk B, Riper H, Ebert DD
Economic Evaluation of an Internet-Based Stress Management Intervention Alongside a Randomized Controlled Trial
JMIR Ment Health 2019;6(5):e10866
URL: https://mental.jmir.org/2019/5/e10866/
doi:10.2196/10866
PMID:31094355

Original Paper

The Importance of User Segmentation for Designing Digital Therapy for Adolescent Mental Health: Findings From Scoping Processes

Theresa Fleming[1,2], PhD; Sally Merry[1], MBChB, MD, FRANZCP; Karolina Stasiak[1], PhD; Sarah Hopkins[1], PhD; Tony Patolo[1], MA; Stacey Ruru[1], MA; Manusiu Latu[1], MPH; Matthew Shepherd[3], PhD; Grant Christie[1], MBChB, MD; Felicity Goodyear-Smith[4], MBChB, MD

[1]Department of Psychological Medicine, University of Auckland, Auckland, New Zealand
[2]Faculty of Health, Victoria University of Wellington, Wellington, New Zealand
[3]Department of Social Work, University of Auckland, Auckland, New Zealand
[4]Department of General Practice & Primary Health Care, University of Auckland, Auckland, New Zealand

Corresponding Author:
Theresa Fleming, PhD
Department of Psychological Medicine
University of Auckland
Private Bag 92019
Auckland, 1142
New Zealand
Phone: 64 9 923 5025
Email: t.fleming@auckland.ac.nz

Abstract

Background: New Zealand youth, especially those of Māori and Pacific descent, have high rates of depression, anxiety, and self-harm, but have low rates of help-seeking from mental health professionals. Apps, computerized therapy, and other digital tools can be effective, highly scalable treatments for anxiety and depression. Co-design processes are often used to foster engagement with end users, but this does not always lead to high levels of engagement.

Objective: We aimed to carry out preliminary scoping to understand adolescents' current internet use and diversity of preferences to inform a planned co-design process for creating digital mental health tools for teenagers.

Methods: Interactive workshops and focus groups were held with young people. Data were analyzed using a general inductive approach.

Results: Participants (N=58) engaged in 2 whānau (extended family) focus groups (n=4 and n=5), 2 school- or community-based focus groups (n=9 each), and 2 workshops (n=11 and n=20). The authors identified 3 overarching themes: (1) Digital mental health tools are unlikely to be successful if they rely solely on youth help-seeking. (2) A single approach is unlikely to appeal to all. Participants had diverse, noncompatible preferences in terms of *look or feel* of an app or digital tool. The authors identified 4 user groups *players or gamers*, *engagers*, *sceptics*, and *straight-talkers*. These groups differed by age and degree of current mental health need and preferred gamified or fun approaches, were open to a range of approaches, were generally disinterested, or preferred direct-to-the-point, serious approaches, respectively. (3) Digital mental health tools should provide an immediate response to a range of different issues and challenges that a young person may face.

Conclusions: Defining the preferences of different groups of users may be important for increasing engagement with digital therapies even within specific population and mental health–need groups. This study demonstrates the importance of scoping possible user needs to inform design processes.

(JMIR Ment Health 2019;6(5):e12656) doi:10.2196/12656

KEYWORDS

gamification; computerized therapy; mental health; adolescent; mobile apps; internet

Introduction

Background

Depression and anxiety, including clinical disorders and subclinical symptoms, are common among young people. New Zealand youth have high rates of mental distress, with rates particularly high among Māori and Pacific girls [1,2]. Most young people do not seek professional help for mental health needs [3], and this *forgone care* is particularly high among indigenous and minority youth [1,4,5].

Digital mental health tools have been shown to be effective and appealing treatments for anxiety and depression [6-8]. This has been demonstrated among different cultural and demographic groups, including Māori and Pacific youth [9,10]. By making interventions available conveniently, without the barriers of access to face-to-face services, it is anticipated that more youth will have access to a variety of treatments more equitably. However, uptake of and adherence to evidence-based digital interventions as implemented have not been strong to date [11,12].

There are many possible ways to improve the population impact of digital interventions, both in the development of tools and in their delivery. Those developing interventions can enhance user engagement through co-design approaches, in which potential users are actively involved in the development process, ensuring that the products or outputs are optimally placed to meet their needs. Other opportunities include the increased use of telepresence, gamification, and contemporary styles and formats [13]. Implementation processes are also important. For example, promotional activities are needed to ensure that potential users and those who might support them are aware of and interested in digital health options. Embedding and integrating digital mental health tools within existing health or education formats or providing integrated clinical support could also increase uptake and use [13].

Co-Design

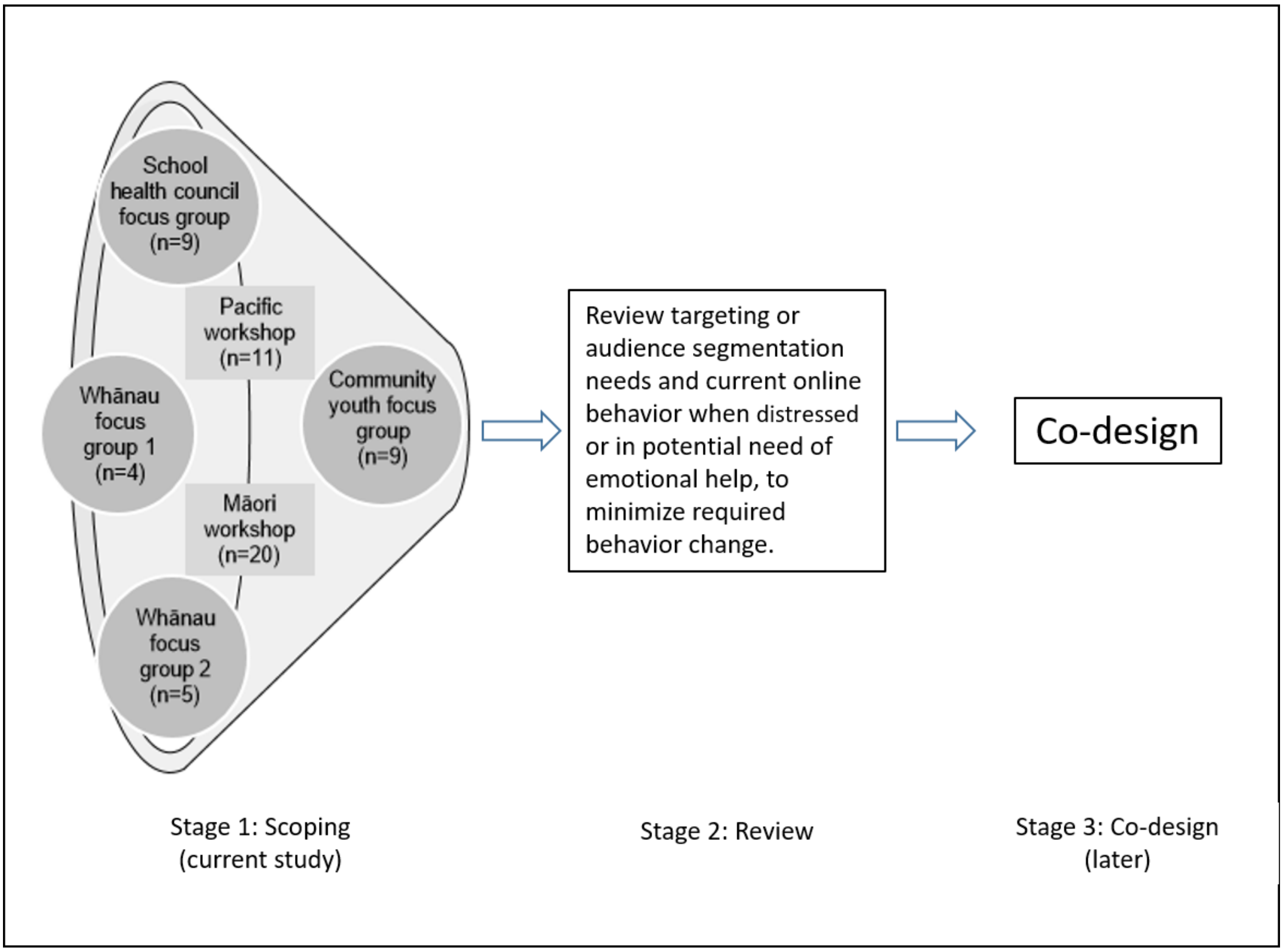
Co-design approaches have been proposed to increase uptake and engagement [14]. In co-design, end users are involved in the development of the intervention through an ongoing democratic partnership with researchers, with a value-based focus on relationships and distribution of power between the *researchers* and the *researched* [15]. Co-design processes can also mean that contextual variables and issues with usability and implementation that may not be obvious to researchers are considered in development. Hence, these processes may increase the acceptability of the end product to the user group. Although there are compelling reasons to involve end users in design, there has been limited evidence to date that consumer involvement in the design of youth mental health interventions improves intervention effectiveness [16]. Furthermore, co-design presents multiple challenges. Co-designers are a subset of the proposed end users and their engagement in the design process can change their preferences. Findings may not generalize to users who were not engaged in development or to those of different ages, cultural groups, and level of need. If end users have diverse or differing needs, there is no guarantee that co-design will enhance impact for all end users. For example, adolescents who are not depressed typically view seeking help and taking action as relatively straightforward, whereas those experiencing depression may find these steps more challenging [17]. Rapidly changing digital technology standards and expectations also mean that findings may not be generalizable over time.

Māori and Pacific peoples form a significant minority of the New Zealand population (approximately 15% and 8%, respectively), are over-represented in higher deprivation settings [18], and are often poorly served by mental health services situated within Western science and health traditions [19]. Māori and Pacific models of mental health and well-being are frequently described as more holistic and family-centered than traditional Western models [19-21]. An explicit focus on underserved communities is important for ensuring relevant services and for supporting equity outcomes.

Aims

We undertook a preliminary scoping phase to inform a planned co-design process, as shown in Figure 1. Our aim was to explore diverse adolescents' preferences to understand the need for specific targeting or audience segmentation and key needs of different populations. We also aimed to explore adolescents' current Web-based behavior when distressed or in potential need of emotional help, to minimize required behavior change.

Figure 1. Scoping using diverse processes to inform co-design.



Methods

Overview

The scoping study took place from December 2016 to April 2017. Focus groups in whānau (extended family), school and community groups, as well as interactive workshops, were planned to engage young people in varied ways and encourage a broad spectrum of feedback from the target audience. Ethics approval was obtained from the Health and Disability Ethics Committee (Ref no 16/NTX/174). Young people were included if they were aged between 12 and 25 years, could provide informed consent, and could participate safely. No young people were excluded. We purposively sampled high proportions of Māori and Pacific young people from varied socioeconomic backgrounds and sought to include participants with personal experience of mental distress by recruiting via community groups and networks that included young people with these characteristics.

Focus Groups

Focus groups of 30- to 60-min duration were held with young people from (1) a school health council, (2) a community youth health organization, and (3) 2 whānau groups. Each whānau focus group included Māori young people who were related to each other, and each took place in a family home. Participants were siblings, cousins, and close family friends aged between 12 and 25 years. Adult family members welcomed the researchers and participants and encouraged participants to take part but were not in the room at the time of the group meeting.

Interactive Workshops

A total of 2 workshops of approximately 4 hours each were run during school holidays. The first was at a community youth center with an existing youth advisory group, predominantly of Pacific descent. The other was at a Māori health organization site with attendees primarily comprising Māori youth group leaders, members of a youth organization, and school students. These workshops were designed to engage Pacific and Māori youth, respectively, and to encourage their participation.

Procedures

Participants were asked about their current behavior and needs regarding online mental health tools, their views on currently available online support options, and their preferences regarding pathways to finding the tool, the modality, the use of gamification, and feedback options. They were given tablets loaded with a range of existing mental health apps and websites to prompt discussion. These apps and websites were readily available in New Zealand, were potentially familiar to young people, and displayed a range of styles or ways of providing digital mental health and well-being support. In particular, we included 4 information, treatment, or support resources funded by the New Zealand government (SPARX, The Lowdown, depression.org and Aunty Dee), an Australian mental health resource (Biteback), and several apps available via the App Store (Smiling Mind, Headspace, and SuperBetter).

In each setting, and particularly in the extended workshops, time was allocated to interactive activities such as icebreakers and open discussions using a cultural process that encourages participants to share experiences and stories, an approach commonly used in Pacific and Māori research [22]. This approach takes into consideration relationships that lay the foundation for most Pacific activities [22]. Pacific authors highlight that maintaining good relationships is important for the success of research as they establish trust between researcher and participants. The process further eliminates the distance between the 2, allowing for flexibility in conversation. The Māori workshop followed culturally appropriate protocol. There was a formal welcome of the researcher guests, with an elder opening the session with a prayer, followed by a formal speech. The guests were then invited to introduce themselves and build relationships with the Māori youth. The elder also closed the workshop according to traditional protocols with prayer. The content of the workshops included discussion, sketching ideas on paper, telling stories and *wall storms* where participants wrote ideas on sticky notes, posted them on a wall, then shared and processed them as a group.

At the end of the sessions, participants completed a written summary reporting their age, ethnicity, gender, and whether they had ever felt low or depressed for more than few days in a row. As is customary in Māori and Pacific research, food was provided as part of the process of engaging with people from the community. Each participant was given a voucher as thanks for taking part.

Data Analysis

Audio recordings from all sessions were transcribed, and facilitators summarized data from sessions that were not recorded due to practicalities (eg, *wall storms* do not lend themselves to audio recording and transcription). Raw data were analyzed and coded in QSR NVivo software. A general inductive approach was used, allowing the main themes to arise from young people's feedback [23,24]. All transcripts were individually coded among several researchers, including Māori and Pacific coders. First, data were clustered together in categories or groupings. Researchers then convened to discuss potentially important themes and clusters and review the emerging themes against the predefined research questions. This process was repeated several times to ensure data were accurately represented and differing opinions settled. This general inductive approach is an appropriate method for scoping general usability preferences as it highlights information that is relatively exploratory [24].

Results

Participants

There were a total of 58 participants drawn from 2 whānau groups (n=4 and n=5), 1 School Health Council group (n=9), 1 Community Youth group (n=9), 1 Pacific workshop (n=11), and 1 Māori workshop (n=20). The majority of participants were female and Māori or Pacific (see Table 1 for demographics). A total of 3 high-level themes emerged.

Table 1. Demographic details of study participants (N=58).

Demographics	Focus groups, n				Interactive workshops, n		Total (N=58), n
	Whānau 1 (n=5)	Whānau 2 (n=4)	School health council (n=9)	Community youth group (n=9)	Pacific (n=11)	Māori (n=20)	
Gender							
Female	3	4	8	7	8	11	41
Male	2	0	1	2	3	5	13
Missing	0	0	0	0	0	4	4
Felt low or depressed for more than a few days in a row							
Yes	1	4	6	8	6	11	36
No	4	0	3	1	5	4	17
Missing	0	0	0	0	0	5	5
Ethnicity							
Māori	1	2	0	1	1	12	17
Pacific	3	0	5	1	8	1	18
NZE[a]	1	2	0	4	1	1	9
Other	0	0	3	1	1	1	6
Missing	0	0	1	2	0	5	8
Age (years)							
<14	1	0	1	0	0	0	2
14-17	0	4	8	1	8	5	26
>17	4	0	0	8	3	10	25
Missing	0	0	0	0	0	5	5

[a]NZE: New Zealand European.

Theme 1: Digital Mental Health Tools are Unlikely to be Successful if They Rely on Youth Help-Seeking

Respondents described a range of strategies they used to deal with distress. Social withdrawal, using drugs and alcohol, and self-harm were common. Other responses included ignoring the distress or using music, prayer, sport, or other activities as a distraction. Most expressed a preference for “keeping personal stuff” to themselves and coping on their own, for example:

> *I don't really tell anyone or people about what I'm going through.* [School Health Council]

> *People probably don’t say anything, just keep it to themselves.* [Whānau group 1]

Participants were concerned and embarrassed about others knowing they were “having a hard time.” Talking distress over with friends and/or family and/or seeking help from a known adult (such as a trusted teacher) were relatively rare, and no participants mentioned seeking help from unknown adults. Despite this, some participants reported that they would post about their distress on social media (such as Facebook or Instagram), on a personal account or anonymously, in the hope that others might understand or reach out to them. For example:

> *You have 8 different accounts and nobody knows who you are. So there’s no stigma attached to saying anything. You can just be truthful.* [Community youth group]

> *Like on Facebook, you can do like feeling, you can put up an emotion to what you’re feeling. So people would be like sad face- feeling down or something…* [Whānau group 1]

Aside from possible posts on social media, no participants spontaneously mentioned that they would seek help or use the internet for mental health needs. They reported that they would feel labeled and stigmatized after accessing mental health websites even if nobody knew:

> *Yeah, you’d feel embarrassed [accessing a mental health website].* [Whānau group 1]

In response to direct questions from the researchers, participants reported that they would seek help only if they were “desperate,” in which case their preference may be posting on social media where friends might respond.

Despite this lack of active help-seeking, participants responded positively to the idea of online support:

> *I think it sounds like a great idea because it’s on mobile and a lot of people are getting phones and stuff…everyone, like especially the teenagers, they’re always on their phone…So it’s cool to have it on their phones.* [School Health Council]

> *...the power that comes from an app is the fact that apps have push notifications and stuff like that, that remind you [to] engage with them…an app would be a good place to have your more interactive parts, and then a website would be a good place to have facts and information and where you can seek help and stuff…* [Community youth group]

These findings were considered alongside the imperative to require realistic behavior change for the use of any digital mental health app or tool. The theme of *Digital mental health tools are unlikely to be successful if they rely on youth help-seeking* was articulated and confirmed by rereading transcripts, reviewing original codes, and discussion among researchers.

Theme 2: A Single Approach is Unlikely to Appeal to All

Importantly, different groups and individuals had diverse, noncompatible preferences in terms of *look or feel* of an app or digital tool. Being relevant and appealing for young people was considered key; pitching information in the right way also mattered. However, what was *the right way* differed among participants. Many young people preferred a crisp, clean look, which was uncomplicated with straightforward, direct, clear, and “straight-to-the-point” help that they would access when they or others were in a crisis. They considered they would not normally look for this kind of help unless they were very worried about someone or were suicidal. In this context, they prioritized relatively serious content; some felt that a game or narrative approach would be off-putting or inappropriate. This view was more frequently expressed by older participants and those who indicated that they had had some experience of distress:

> *That person has literally hit the point where they can’t reach out anymore. They need something that engages with them instantly [rather than is playful or indirect].* [Community youth group]
>
> *If it was like gaming aspects in there, like characters, like moving up a level or something, [it’s not right] because it is quite a serious matter. So yeah, take it seriously.* [School Health Council]
>
> *I’d go straight to the point kind of stuff. Like I’d want to know bullet point, then another bullet point.* [Whānau group 1]

In contrast, other participants considered that a *gamey*, fun, or entertaining environment could engage people who might not otherwise recognize that they need help or be seeking help. A nonthreatening, normalizing, nonpathologizing approach was seen as key, and participants suggested focusing on identity, everyday problems, well-being, strengths, and inspiration. These were often younger participants with less experience of mental distress:

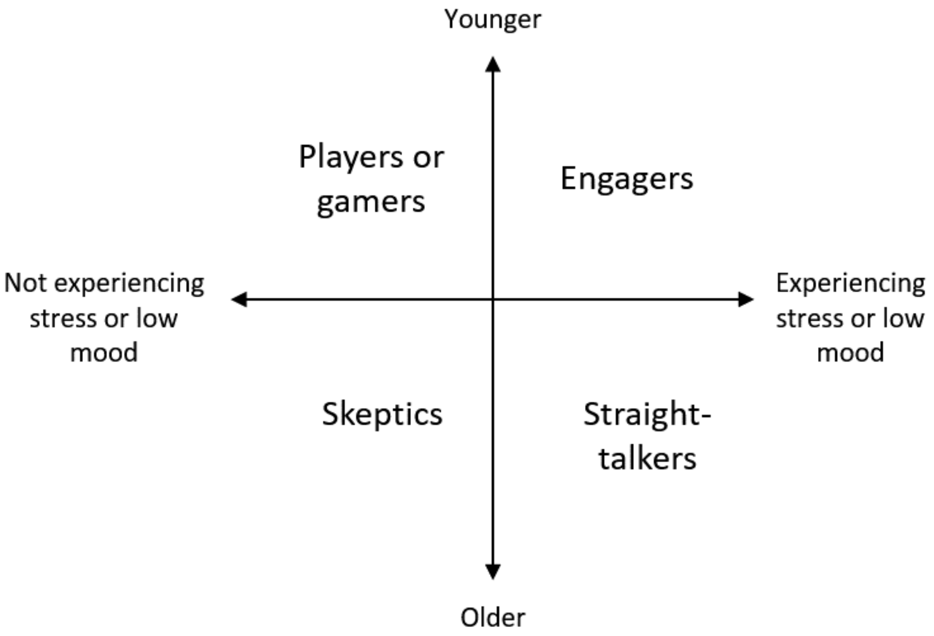
> *I do like the idea of a game. Yeah, that is a cool idea.* [Whānau group 2]
>
> *Like it’s actually engaging but it’s also teaching you stuff.* [Whānau group 2]
>
> *And yeah, try and make it as a game. Yeah, ‘cos it’s fun.* [School Health Council]

These findings were considered alongside the research needs. The theme of *a single approach unlikely to appeal to all* was articulated and confirmed by rereading transcripts and reviewing original codes. Next, we clustered user preferences. A total of 4 major groupings or categories emerged:

1. Those who were more interested in a digital intervention to support their mental well-being if it was *fun* to do. Many reported that they would engage in an intervention that was storified or gamified. These participants were often younger adolescents who were not feeling stressed out or down at the time of the study. We identified them as *Players or gamers.*
2. Adolescents who were interested in digital interventions that were targeted directly to support their mental well-being, interactive and potentially gamified. Again, these participants were often younger adolescents, but they generally had more experience of feeling stressed or down than *Players*. We categorized them as *Engagers.*
3. Young people who were not personally interested in digital interventions for mental health. These participants felt no need for such interventions and were skeptical about their value, expressing more interest in social media and interacting with their friends. This group generally comprised older adolescents who were not feeling stressed out or down at the time of the study. We identified this group as *Skeptics*.
4. Finally, a cluster of young people expressed interest in *straight to the point* digital interventions for mental well-being. They considered gamification off-putting and indicated a preference for something simple with a clean, clear design that conveyed the message *straight up*. This cluster mainly comprised older adolescents experiencing stress or low mood. We categorized this group as *Straight-talkers*.

We proposed that these clusters can be represented in a taxonomy as shown in Figure 2.

Figure 2. Taxonomy of adolescents engaging with digital mental health tools.



Theme 3: Help Here and Now, With Multiple Challenges

Despite differences in preferred style, there was a high level of consensus that young people preferred content that made them feel good and helped them cope with hard times, with impact in the *here and now* rather than in the future. Participants also reported that they wanted freedom of choice, including the ability to pick and choose interventions and would want help with a variety of challenges or needs such as bullying, relationship problems, self-harm, suicide, and other mental health needs:

> *Today [it] is all about choice, really.* [Community youth group]
>
> *Yeah I think apps that had different components, so maybe like a combination of meditation and information or interactive kind of things* [Whānau group 2]

Discussion

Principal Findings

Participants reported that they did not actively seek help if down or distressed unless they were very concerned, for example, if there might be a risk of suicide. If they did use the internet when feeling down or distressed, this consisted of posting on a social media site where friends might understand or respond. Despite this lack of active help seeking, participants responded positively to the idea of online support. There were diverse preferences for the look and style of such interventions. Older adolescents and those experiencing distress were interested in direct, to-the-point information and considered that gamification might be trivializing. They wanted tools that connected to social support in some way. Some, especially younger, participants were more interested in game-like and playful approaches that might have a more universal appeal and were less directly mental health–focused. These differences in preferences highlight the challenge in trying to *be all things to all people* and the need for targeted approaches or *audience segmentation* in co-design and marketing terms. Finally, participants highlighted the need for immediate help with multiple challenges.

Comparisons With Previous Work

These results suggest that improving youth mental health via specific digital interventions has promise. However, acknowledging young people's lack of help-seeking and diverse preferences is critical to the successful implementation of digital mental health interventions.

Our finding that participants were unlikely to seek help for mental health needs online is consistent with previous literature [25]. There are major challenges reaching groups that have low rates of help-seeking if approaches rely on users proactively and individually seeking help. Young people are generally poor at seeking help for emotional distress for myriad reasons including time limitations, privacy concerns, financial reasons, and skepticism about treatment effectiveness [26]. Unsurprisingly, those from high deprivation neighborhoods find it hardest to access health care [27]. To reach those with high needs, online interventions may need to be promoted or actively introduced, for example, through routine incorporation into established systems that young people are exposed to, such as school health checks or the school health curriculum. For those not attending school or where a school-based approach is less likely to hold appeal, an approach that utilizes commonly used media may be more appropriate. For example, many young people currently use *Messenger*; a chat-bot using this conversational channel may make an intervention more appealing and accessible. Previous research has highlighted that young people do look for health information online [28,29]. Hence, search engine advertisements and websites providing information could be used to help encourage young people to seek help, whether online or face-to-face.

Our finding that participants had diverse and often conflicting preferences, even within this relatively specific demographic, shows that efforts to deliver digital interventions risk failure if developers try to be all things to all people. In fact, if we were to *average* all results, we might be inclined to suggest that an intervention needed to be *a little bit playful*, whereas in reality, results were dichotomous. This finding should come as no surprise as a similar phenomenon is apparent among those accessing the internet for other reasons. For example, Bartle, long ago proposed a taxonomy of *gamers* with differing views about what appeals in computer games [30]. He proposed 4 categories of gamers: killers, explorers, socializers, and collectors. This idea has since been expanded [31]. In many current games, one can see specific designs to appeal to the varying populations of likely users. Our finding of broad clusters of users and our proposed taxonomy of young people who use digital interventions to support their mental health is consistent with these concepts of audience segmentation.

Such analyses can be used to inform the design of digital interventions to appeal to specific groups. For example, interventions that aim to prevent emotional health problems might be delivered to young groups using gamified or storified approaches. Designing for young people who are seeking help for established problems might require a more direct approach.

Implications

The results of this study suggest that online approaches to improve the mental health of young people, particularly Māori and Pacific young people in New Zealand, should consider the following needs:

- Direct and clear information and immediate access to support for those who are seeking help in a crisis;
- Interventions targeting those posting about their distress on social media (eg, this can include social media platforms providing direct-to-user messages or alerts to services, based on user-generated content);
- Approaches that engage young people before their needs are high and before they would be seeking help in current contexts. This might be done via health promotion and marketing to attempt to increase interest and reduce stigma [32], increasing the attractiveness and appeal of interventions, or mandated systems that target groups are exposed to (eg, universal screening with interventions via health or education systems).

The wide range of opinions reported by our participants highlights that digital interventions that aim to appeal to all young people risk failure and that more targeted strategies may have more promise. Should future research support our taxonomy of Players or gamers, Engagers, Skeptics, and Straight-talkers, we should expect a range of youth-oriented mental health tools, with some serious and direct and others more playful or gamified. Segmentation requires a greater investment of resources than a single approach that reaches all. In other areas, however, from games to news sites, online tools are diverse, a feature considered essential for engagement. Indeed, a single approach that aims to reach all and in doing so appeals to none or few cannot be classed a success. Increased diversity and targeting of digital mental health options should be explored as an approach to improve engagement.

The identification of conflicting user preferences also highlights the value of scoping user needs and preferences *before* establishing co-design processes. For example, without identifying varied needs, we may have developed a mixed group of co-designers pulling in competing directions or co-operating to produce an outcome not particularly appealing to any.

Finally, certain characteristics appear important across all groups, including the provision of choice. Young people wanted to be able to pick and choose those interventions that suited them. In the same way that providing a *menu of options* is an important component of brief face-to-face interventions, it will be important for digital interventions to provide users with the experience of being in control of the process [33].

Limitations

This is a small and very specific sample. We used a purposive sample that included indigenous Māori and Pacific youth who provided information about their experiences of digital resources; few studies exist that have sought to capture this information. Although useful in providing needed information about this group, our focus on Māori and Pacific youth in New Zealand and small group sizes limits the generalizability of findings to other demographics. Future research should scope diversity of user preferences within other population and health-need groups.

Conclusions

This study demonstrates the importance of scoping and consulting before designing. Youth have varied preferences and it is essential to decide on a focused approach and then co-design with adolescents in a particular group if engagement and implementation are to succeed. The need for varied interfaces is clear.

The use of technology to increase the delivery of psychological interventions to support the mental health of young people remains a promising approach, but successful implementation will require a greater degree of sophistication in design and rollout than has been demonstrated to date.

Acknowledgments

The authors would like to sincerely thank all the young people who generously gave their time and energy, without which this project could not have occurred. The authors also thank Dr Kahu McClintock for her assistance with Māori engagement and workshops. This work forms part of the HABITs project, funded by the New Zealand National Science Challenge–A Better Start/E Tipu e Rea. The authors wish to acknowledge and thank the HABITs investigators and collaborators and funders. They also sincerely thank Kylie Sutcliffe, Victoria University of Wellington, for her editing assistance.

Conflicts of Interest

The intellectual property for SPARX is owned by Uniservices at the University of Auckland, but if income was generated from SPARX, SM, TF, KS, and MS would stand to gain financially.

References

1. Teevale T, Lee AC, Tiatia-Seath J, Clark TC, Denny S, Bullen P, et al. Risk and protective factors for suicidal behaviors among Pacific youth in New Zealand. Crisis 2016 Sep;37(5):335-346. [doi: 10.1027/0227-5910/a000396] [Medline: 27278568]
2. Clark TC, Robinson E, Crengle S, Fleming T, Ameratunga S, Denny SJ, et al. Risk and protective factors for suicide attempt among Indigenous Māori Youth in New Zealand: the role of family connection. J Aborig Health 2011;7(1):16-31. [doi: 10.18357/ijih71201112350]
3. Peni B, Day K, Orr M. What Pacific people think of online mental health information. In: Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management. Darlinghurst, Australia: Australian Computer Society, Inc; 2014 Presented at: HIKM '14; January 20-23, 2014; Auckland, New Zealand p. 51-58.
4. Denny S, Farrant B, Cosgriff J, Harte M, Cameron T, Johnson R, et al. Forgone health care among secondary school students in New Zealand. J Prim Health Care 2013 Mar 01;5(1):11-18. [Medline: 23457690]
5. Mariu KR, Merry SN, Robinson EM, Watson PD. Seeking professional help for mental health problems, among New Zealand secondary school students. Clin Child Psychol Psychiatry 2012 Apr;17(2):284-297. [doi: 10.1177/1359104511404176] [Medline: 21852315]
6. Stasiak K, Fleming T, Lucassen MF, Shepherd MJ, Whittaker R, Merry SN. Computer-based and online therapy for depression and anxiety in children and adolescents. J Child Adolesc Psychopharmacol 2016 Apr;26(3):235-245. [doi: 10.1089/cap.2015.0029] [Medline: 26465266]
7. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
8. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. Brain Sci 2017 Nov 12;7(11):12 [FREE Full text] [doi: 10.3390/brainsci7110150] [Medline: 29137156]
9. Shepherd M, Fleming T, Lucassen M, Stasiak K, Lambie I, Merry SN. The design and relevance of a computerized gamified depression therapy program for indigenous Māori adolescents. JMIR Serious Games 2015;3(1):e1 [FREE Full text] [doi: 10.2196/games.3804] [Medline: 25736225]
10. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
11. Loiselle CG, Ahmed S. Is connected health contributing to a healthier population? J Med Internet Res 2017 Nov 10;19(11):e386 [FREE Full text] [doi: 10.2196/jmir.8309] [Medline: 29127077]
12. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
13. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [Medline: 27148094]
14. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Young and Well Cooperative Research Centre. 2012. Participatory design of evidence-based online youth mental health promotion, intervention and treatment URL: https://www.westernsydney.edu.au/__data/assets/pdf_file/0005/476330/Young_and_Well_CRC_IM_PD_Guide.pdf [accessed 2019-04-17] [WebCite Cache ID 77h75j9R9]
15. Goodyear-Smith F. Collective enquiry and reflective action in research: towards a clarification of the terminology. Fam Pract 2017 Dec 01;34(3):268-271. [doi: 10.1093/fampra/cmw098] [Medline: 27629568]
16. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
17. Wilson CJ, Deane FP. Help-negation and suicidal ideation: the role of depression, anxiety and hopelessness. J Youth Adolesc 2010 Mar;39(3):291-305. [doi: 10.1007/s10964-009-9487-8] [Medline: 19957103]
18. Atkinson J, Salmond C, Crampton P. Department of Public Health, University of Otago. 2014. NZDep2013 Index of Deprivation URL: https://www.otago.ac.nz/wellington/otago069936.pdf [accessed 2019-04-17] [WebCite Cache ID 77h7HSCSf]
19. Durie M. Maori Ora: The Dynamics of Māori Health. Auckland: Oxford University Press; 2001.
20. Pere R. Concepts and Learning in the Māori Tradition. Hamilton: University of Waikato, Department of Sociology; 1982.

21. Tamasese K, Peteru C, Waldegrave C, Bush A. Ole Taeao Afua, the new morning: a qualitative investigation into Samoan perspectives on mental health and culturally appropriate services. Aust NZ J Psychiatry 2016 Jun 26;39(4):300-309. [doi: 10.1080/j.1440-1614.2005.01572.x]
22. Vaioleti TM. Talanoa research methodology: a developing position on Pacific research. WJE 2006;12:21-34 [FREE Full text]
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
24. Thomas DR. A general Inductive approach for analyzing qualitative evaluation data. Am J Med Eval 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]
25. Utter J, Lucassen M, Denny S, Fleming T, Peiris-John R, Clark T. Using the Internet to access health-related information: results from a nationally representative sample of New Zealand secondary school students. Int J Adolesc Med Health 2017 Nov 23:23 Epub ahead of print (forthcoming). [doi: 10.1515/ijamh-2017-0096] [Medline: 29168960]
26. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
27. Clark T, Fleming T, Bullen P, Denny S, Crengle S, Dyson B, et al. The University of Auckland. 2013. Youth'12 Overview: The health and wellbeing of New Zealand secondary school students in 2012 URL: https://cdn.auckland.ac.nz/assets/fmhs/faculty/ahrg/docs/2012-overview.pdf [accessed 2019-04-17] [WebCite Cache ID 77h7gyck3]
28. Utter J, Lucassen M, Denny S, Fleming T, Peiris-John R, Clark T. Using the internet to access health-related information: results from a nationally representative sample of New Zealand secondary school students. Int J Adolesc Med Health 2017 Nov 23 Epub ahead of print(forthcoming). [doi: 10.1515/ijamh-2017-0096] [Medline: 29168960]
29. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, health and technology: a national survey. Media Communication 2016 Jun;4(3). [doi: 10.17645/mac.v4i3.515]
30. Bartle R. MUD. Hearts, clubs, diamonds, spades: Players who suit MUDS URL: http://mud.co.uk/richard/hcds.htm [accessed 2018-10-13] [WebCite Cache ID 77fRnsLt0]
31. Stewart B. Gamasutra. Personality And Play Styles: A Unified Model URL: https://www.gamasutra.com/view/feature/6474/personality_and_play_styles_a_.php?print=1 [accessed 2018-10-13] [WebCite Cache ID 77fS1C15k]
32. Clement S, Lassman F, Barley E, Evans-Lacko S, Williams P, Yamaguchi S, et al. Mass media interventions for reducing mental health-related stigma. Cochrane Database Syst Rev 2013;7:CD009453. [doi: 10.1002/14651858.CD009453.pub2] [Medline: 23881731]
33. Brief intervention for hazardous and harmful drinking : a manual for use in primary care. Geneva: World Health Organization; 2001 Oct 13. URL: http://www.who.int/iris/handle/10665/67210 [WebCite Cache ID 77fRhNY2p]

Abbreviations

HABITs: Health Advances through Behavioural Intervention Technology
NZE: New Zealand European
SPARX: Smart Positive and Realistic X-Factor Thoughts

Edited by G Eysenbach; submitted 07.11.18; peer-reviewed by E Sezgin, A Benis; comments to author 13.12.18; revised version received 07.03.19; accepted 24.03.19; published 08.05.19.

Please cite as:
Fleming T, Merry S, Stasiak K, Hopkins S, Patolo T, Ruru S, Latu M, Shepherd M, Christie G, Goodyear-Smith F
The Importance of User Segmentation for Designing Digital Therapy for Adolescent Mental Health: Findings From Scoping Processes
JMIR Ment Health 2019;6(5):e12656
URL: https://mental.jmir.org/2019/5/e12656/
doi:10.2196/12656
PMID:31066705

Original Paper

Association Between Improvement in Baseline Mood and Long-Term Use of a Mindfulness and Meditation App: Observational Study

Argus J Athanas[1*], BSc; Jamison M McCorrison[2,3*], BSc; Susan Smalley[4], PhD; Jamie Price[5], JD; Jim Grady[5], BA; Paul Wehner[6], BA; Julie Campistron[5], MBA; Nicholas J Schork[1,3,7,8], PhD

[1]Department of Biomedical Informatics, University of California San Diego, San Diego, CA, United States
[2]Department of Bioinformatics and Systems Biology, University of California San Diego, San Diego, CA, United States
[3]J Craig Venter Institute, San Diego, CA, United States
[4]Department of Psychiatry, University of California Los Angeles, Los Angeles, CA, United States
[5]Stop, Breathe & Think, Los Angeles, CA, United States
[6]Positive Place Inc, The Villages, FL, United States
[7]The Translational Genomics Research Institute (TGen), Department of Quantitative Medicine, Phoenix, AZ, United States
[8]The City of Hope/Translational Genomics Research Institute IMPACT Center, Duarte, CA, United States
[*]these authors contributed equally

Corresponding Author:
Nicholas J Schork, PhD
The Translational Genomics Research Institute (TGen)
Department of Quantitative Medicine
445 N 5th St
Phoenix, AZ, 85004
United States
Phone: 1 (602) 343 8400
Email: nschork@tgen.org

Related Article:

This is a corrected version. See correction statement: https://mental.jmir.org/2021/6/e28132/

Abstract

Background: The use of smartphone apps to monitor and deliver health care guidance and interventions has received considerable attention recently, particularly with regard to behavioral disorders, stress relief, negative emotional state, and poor mood in general. Unfortunately, there is little research investigating the long-term and repeated effects of apps meant to impact mood and emotional state.

Objective: We aimed to investigate the effects of both immediate point-of-intervention and long-term use (ie, at least 10 engagements) of a guided meditation and mindfulness smartphone app on users' emotional states. Data were collected from users of a mobile phone app developed by the company Stop, Breathe & Think (SBT) for achieving emotional wellness. To explore the long-term effects, we assessed changes in the users' basal emotional state before they completed an activity (eg, a guided meditation). We also assessed the immediate effects of the app on users' emotional states from preactivity to postactivity.

Methods: The SBT app collects information on the emotional state of the user before and after engagement in one or several mediation and mindfulness activities. These activities are recommended and provided by the app based on user input. We considered data on over 120,000 users of the app who collectively engaged in over 5.5 million sessions with the app during an approximate 2-year period. We focused our analysis on users who had at least 10 engagements with the app over an average of 6 months. We explored the changes in the emotional well-being of individuals with different emotional states at the time of their initial engagement with the app using mixed-effects models. In the process, we compared 2 different methods of classifying emotional states: (1) an expert-defined a priori mood classification and (2) an empirically driven cluster-based classification.

Results: We found that among long-term users of the app, there was an association between the length of use and a positive change in basal emotional state (4% positive mood increase on a 2-point scale every 10 sessions). We also found that individuals who were anxious or depressed tended to have a favorable long-term emotional transition (eg, from a sad emotional state to a happier emotional state) after using the app for an extended period (the odds ratio for achieving a positive emotional state was 3.2 and 6.2 for anxious and depressed individuals, respectively, compared with users with fewer sessions).

Conclusions: Our analyses provide evidence for an association between both immediate and long-term use of an app providing guided meditations and improvements in the emotional state.

(JMIR Ment Health 2019;6(5):e12617) doi:10.2196/12617

KEYWORDS

mental health; smartphone; emotional well-being; mindfulness

Introduction

Background

Behavioral conditions, neuropsychiatric diseases, and poor general mental health are seen as major contributors to morbidity, mortality, and lost productivity on a global scale. However, these factors are often overlooked in discussions about the current state of health care, which tend to focus on physical well-being [1]. Many studies suggest that mental health can play a large role in physical health, recovery from disease, and ultimately productivity and, therefore, should receive greater attention [2-4]. Unfortunately, there are serious questions about how mental health can be promoted and, in instances when it is called for, how relevant interventions can be prescribed and deployed efficiently in a cost-effective manner [5-7]. This is especially true given the number of people who may actually benefit from such interventions [8]. In light of this, there is enthusiasm for the development of smartphone apps that can not only monitor an individual's health—both physical and mental—but also deliver content designed to help coach them through difficult times or provide a needed intervention. In fact, many smartphone apps have been developed, or are under development, to aid in health care via, for example, image-based diagnostics, glucose monitoring for diabetes, and physical fitness promotion [9,10]. For mental health management and intervention, there is growing enthusiasm for the development of smartphone platforms that provide guidance on mindfulness and meditation as a way of relieving stress and promoting mental health and well-being. Many of the resulting platforms have been or are undergoing testing in clinical studies [11-15].

The use of mobile phone apps in combating or mediating behavioral conditions, stress, negative emotional states, and elevating mood is also consistent with directions that public health and regulatory officials are considering. In fact, evidence is mounting from clinical trials showing that smartphone apps can be effective in a variety of settings. Agencies such as the US Food and Drug Administration (FDA) have created, and in instances passed, legislation allowing the filing and approval of mobile health apps as approved health technologies on the same level as in vitro diagnostics and drugs. Pear Therapeutics was one of the first companies to have a smartphone app for addiction approved for use by the FDA in 2016 [16]. Many other commercial and academic groups are developing smartphone apps for a wide variety of conditions that go beyond the simple direct-to-consumer market by seeking regulatory approval for their use in clinical contexts [17-19]. Unfortunately, not enough time has elapsed since the introduction of smartphone-based intervention apps to provide insight into their long-term repeated effects as well as their effects in real-world settings (ie, outside of clinical trials) [20-22].

Objectives

Stop, Breathe & Think (SBT) has developed a smartphone app that provides guided meditations and mindfulness activities to promote self-awareness coaching to interested users. As noted, mindfulness and meditation have been shown to improve affect and mood and promote healthy thought patterns [23,24]. The SBT app prompts users before and after they are guided through meditation and mindfulness activities to provide an emotional, mental, and physical *check-in*, thereby allowing an assessment of an individual user's emotional state and mood pre- versus postactivity in real time. As repeated uses of the app by SBT users are archived, longitudinal information on its users with regard to their long-term engagement with the app is retained. This allows further analysis of the influence of repeated engagements with the app on an individual user's basal mood over time in real-world settings. We pursued such an analysis using data from SBT users who had at least 10 engagements with the app. The SBT app allows users to choose from more than 100 unique emotions to reflect their emotional state at the time they use the app. These emotions cover a range of human emotions including anger, remorse, anxiety, calmness, and enthusiasm. Users are guided through meditations that they can choose from based on an algorithm developed by SBT. We focused our analyses on the *baseline* (or *basal*) emotional state of a user, before he or she engaged in a guided meditation or mindfulness activity and were primarily interested in the long-term and repeated use effects of the SBT app on this baseline emotional state. Essentially, we wanted to ask the question if the continued use of the app lifted the spirits of the user over time. We were particularly interested in users who tended to pick emotions associated with depression and anxiety when engaging with the app before meditating.

Methods

The Basic Stop, Breathe & Think App

The SBT app is a multiplatform (ie, iOS, Android, and Alexa) app designed to guide users through meditations and mindfulness activities to alleviate stress, anxiety, and depression and improve the sense of well-being. Upon opening the app, a user can participate in an optional 10-second reflection period. After this optional reflection period, users describe their current mood, emotional state, and physical health by choosing from a number of emotions; the SBT app then provides suggestions for specific meditation and mindfulness activities. The user can choose from among the suggested activities after being asked to endorse up to 5 different characterizations of their mood and emotional state. A user can choose not to provide any input regarding their mood, emotional state, and physical health and simply engage in an activity.

Figure 1. Stop, Breathe & Think user interface and stages of interaction with the app. Users are provided several ways in which they can record their current emotional state both pre- and postactivity. These emotional check-ins are optional, but the intuitive and simple selection process makes it easy for most users to enter at least some emotional status information.

Step 0. Begin session with optional 10-second reflection.

Step 1. Select physical and then mental state on 5-point scale. Option to skip.

Step 2. Select up to 5 emotions from five categories. Option to skip.

Step 3. Select an activity, and follow prompts.

Step 4. Option to select another activity and return to step 3 or continue

Step 5. Optional mental, physical, and emotional postsession check in

Figure 1 provides a schematic of an individual session and the corresponding points where user information is collected.

It should be understood that all information collected with the SBT app is volunteered by users as stated and defined in the SBT user licensing agreement and privacy policy. In addition, for purposes of our data analyses, all the data we obtained from SBT were anonymized and put into a Health Insurance Portability and Accountability Act (HIPAA)-compliant format such that users could not be reidentified. Functionality and delivery of the SBT app and service varies from device and platform implementation (eg, Alexa, Android, and Web browser). Therefore, to avoid batch effects, we focused on users who were exclusively on an iOS platform and started using the app after SBT provided its last major version of the app (05/01/2016). Users had to have completed at least 10 sessions or engagements with the app, with a minimum of 6 of those sessions including pre- and postactivity emotion selections. The SBT app content is in English and to avoid translation errors and alternative interpretations of the language used in the SBT app, we restricted our analyses to individuals from native English-speaking countries: the United States, United Kingdom, Canada, and Australia. An additional filter was used, restricting users' ages to between 12 and 100 years.

Emotional Check-ins Pre- and Postactivity Score

The SBT app allows the user to endorse between 1 and 5 emotional states out of a possible 115, before and after engagement in a guided meditation or mindfulness activity (or series of activities if they choose to engage in more than 1 activity during a session). This emotional *check-in* involves selecting an initial emoticon and then choosing from a list of emotions within subgroups of terms that closest characterize the user's current emotional state. These 115 emotions were chosen for the app based on internal SBT research and user requests. All emotions were classified as positive, neutral, or negative and given corresponding scores of 1, 0, and −1, respectively. All emotions and their corresponding scores are provided in Multimedia Appendix 1. As users can select up to 5 emotions, an average emotional score was calculated for both pre- and postactivity and standardized to a range from −1 (all negative emotions) to 1 (all positive emotions). Our analysis explored (1) trends in the preactivity emotional score over repeated uses of the app while accounting for the covariates as well as serial correlation between sessions and (2) trends in changes of the emotional scores before and after an activity over repeated uses of the app.

Clustering of Emotions

In addition to treating the preactivity emotion scores and changes in emotion scores pre- and postactivity as dependent variables and time, sex, and age covariates as independent variables, we also explored the patterns among the emotion endorsements to see if there was evidence for obvious clusters of emotions that could reflect the same general emotional state. We leveraged principal coordinates analysis (PCoA) and the nonsupervised clustering technique, *Partitioning Around Medoids* (PAM), for these analyses [25]. We pursued these analyses as it is arguable that some users may see a subset of the emotions as synonymous and hence only choose one among many possible choices to describe their emotional state at the time to avoid redundancy, whereas other users might see those same subsets of emotions as complementary and reflecting different aspects of their mood. In addition, other users may preferentially select emotions based on their location in the selection list or choose a set of rare emotions that are infrequently selected by other users to differentiate their emotions.

The distance between the emotions was calculated using the Bray-Curtis distance measure [26]. To determine the optimal number of nonsupervised emotion clusters in 2-dimensional PCoA component space, we selected the number of clusters with the largest silhouette score. Once we identified the optimal number of clusters, emotions were then assigned to one of the identified clusters.

An individual's emotional status was also summarized in terms of the relative *distances* (using the Euclidean distance measure) between pre- and postactivity states. The distances between an individual's emotional status and the medoid of the closest associated emotion cluster were calculated as well. Emotions were labeled with clinical categories, associating each of them with either anxiety, depression, anger, or happiness (Multimedia Appendix 1). Ultimately, using distances between emotional states and emotional clusters allowed us to build models relating the number of times users engaged with the app to gross changes in emotional states defined by the emotion clusters.

Statistical Analyses to Identify Long-Term Changes in Emotional State

To assess the effects of the continued use of the app on the preactivity emotional state, we used Linear Mixed-Effects (LME) models and Generalized Linear Models (GLMs) as implemented in the lme4 package in R [27]. These analysis techniques can accommodate serial correlations among emotions over time and also account for both fixed (eg, sex) and random effects (eg, variation in preactivity emotional state or the degree to which use of the app changed the preactivity emotional state over time). We pursued different analyses to evaluate changes in the preactivity emotional state over time, including a model that considered the effect of the emotional states possessed by individuals at their first engagement with the app. These analyses considered both the emotion scores as the dependent variables as well as the use of the emotions as defined by the cluster analysis clinical labels as dependent variables. We also tested the effect of repeated uses of the app on the change in the emotional state pre- to postactivity by treating the ratio of pre- to postemotion score as a dependent variable.

We included several covariates in our analyses and tested them for their effects on the emotional state: session index (ie, 1 as the first use and 2 as the second use—which captures the repeated use of the app), gender, age, country of origin, subscription status, and whether the user remained anonymous (ie, did not fill out information in his or her account—which may indicate a fake or disengaged user). As there is large variability in the number of completed sessions and the distribution of the number of uses of the app per individual has an extreme right skew, we applied a $\log_{10}$ transformation to the session index variable. This transformation markedly improved the normality of the session index as a variable (data not shown). LME models were fit, and the features associated with the preactivity emotional state as the dependent variable were selected using a forward stepwise selection procedure based on the Akaike Information Criteria. Similar models were fit with the pre- to postactivity emotional state ratio as the dependent variable. GLMs were fit to the data when changes in emotion categories (ie, based on clinical or cluster analysis labels) were taken as the dependent variable.

Results

Defining the Dataset

After all the duration, quality, platform, and country filters were applied, 13,393 users remained (10,082 females, 2187 males, and 1124 undeclared sex). The average age of the users was 32.3 (SD 13.5) years, with 31.7 (SD 13.3) years for females, 34.6 (SD 13.4) years for males, and 33.3 (SD 15.0) years for undeclared participants. Collectively, the users completed 569,961 sessions with the app, with 302,514 of these sessions having emotional check-in data, with an average of 42.6 sessions and 22.6 emotional check-ins per user. Multimedia Appendix 2 provides a histogram depicting the distribution of the length of time users engaged with the app. Multimedia Appendix 3 shows average period between app uses given the total length of engagement for users.

Cluster Analysis of the Emotions

The use of the silhouette scores based on the PCoA and PAM analyses suggested that there were likely 8 clusters of emotions [28]. As noted, the relative distances between pre- and postactivity emotional states and the distances between each user's emotional state and the closest associated emotion cluster were calculated. In addition, each of the 115 emotions that could be endorsed was assigned to one of the emotion clusters (see Multimedia Appendix 1). Using these cluster labels, we calculated the mean orientation of each cluster and the relative distance of each individual's emotional scores both pre- and postactivity from these means. These distances were compared with the other emotion scores we calculated and were highly correlated with them (Figure 2). Figure 3 provides a graphical depiction of the results of the clustering using the first 2 principal coordinates obtained from our analyses.

Figure 2. Average emotional score versus cluster centroid distances correlation matrix represented as a heat map. As an example for interpreting the numbers in the matrix, a −0.90 correlation between the preactivity emotion score (x-axis Average Pre Emo Score label) and positivity cluster (y-axis Dist positivity label) shows that users who score higher on the preactivity emotional score had a shorter distance of their selected emotions to the centroid of the positive emotion cluster. Note that labels with Dist reflect distance measures derived from the cluster analyses (eg, Dist Anxiety reflects the distance of a user's emotional score from the anxiety cluster mean) and Emo reflects a specified emotional cluster.

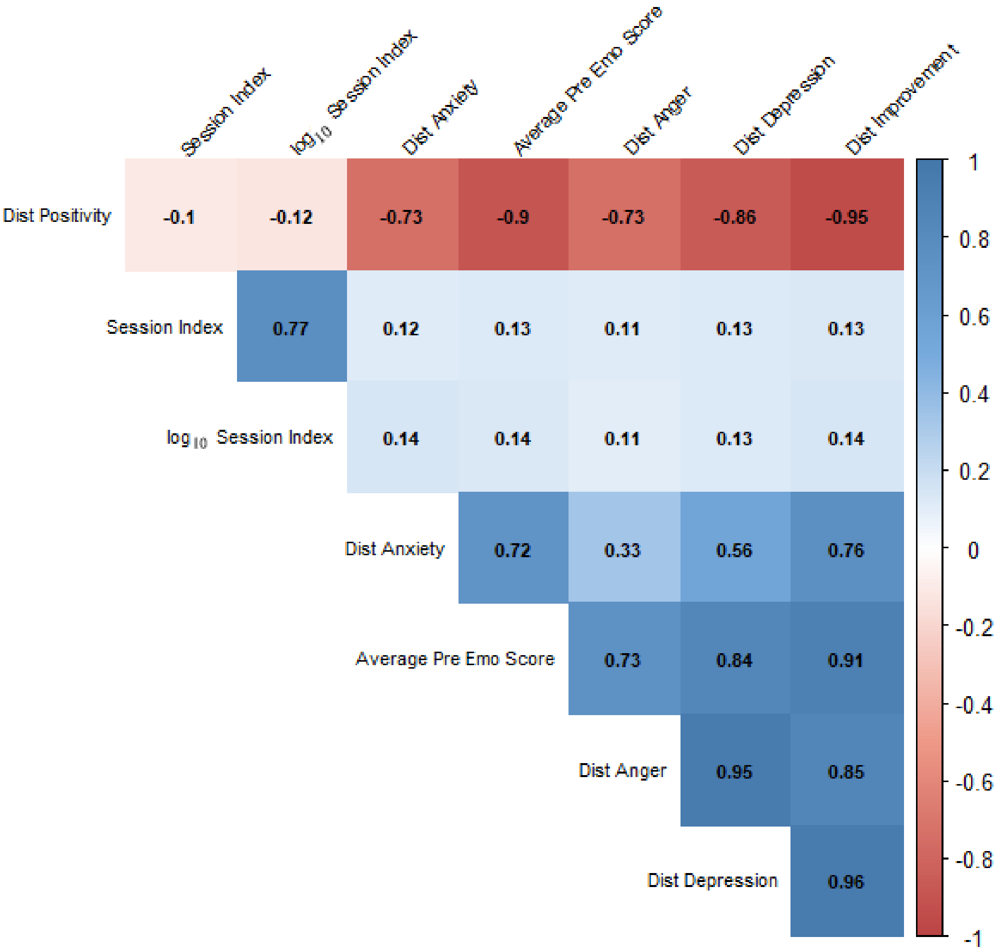
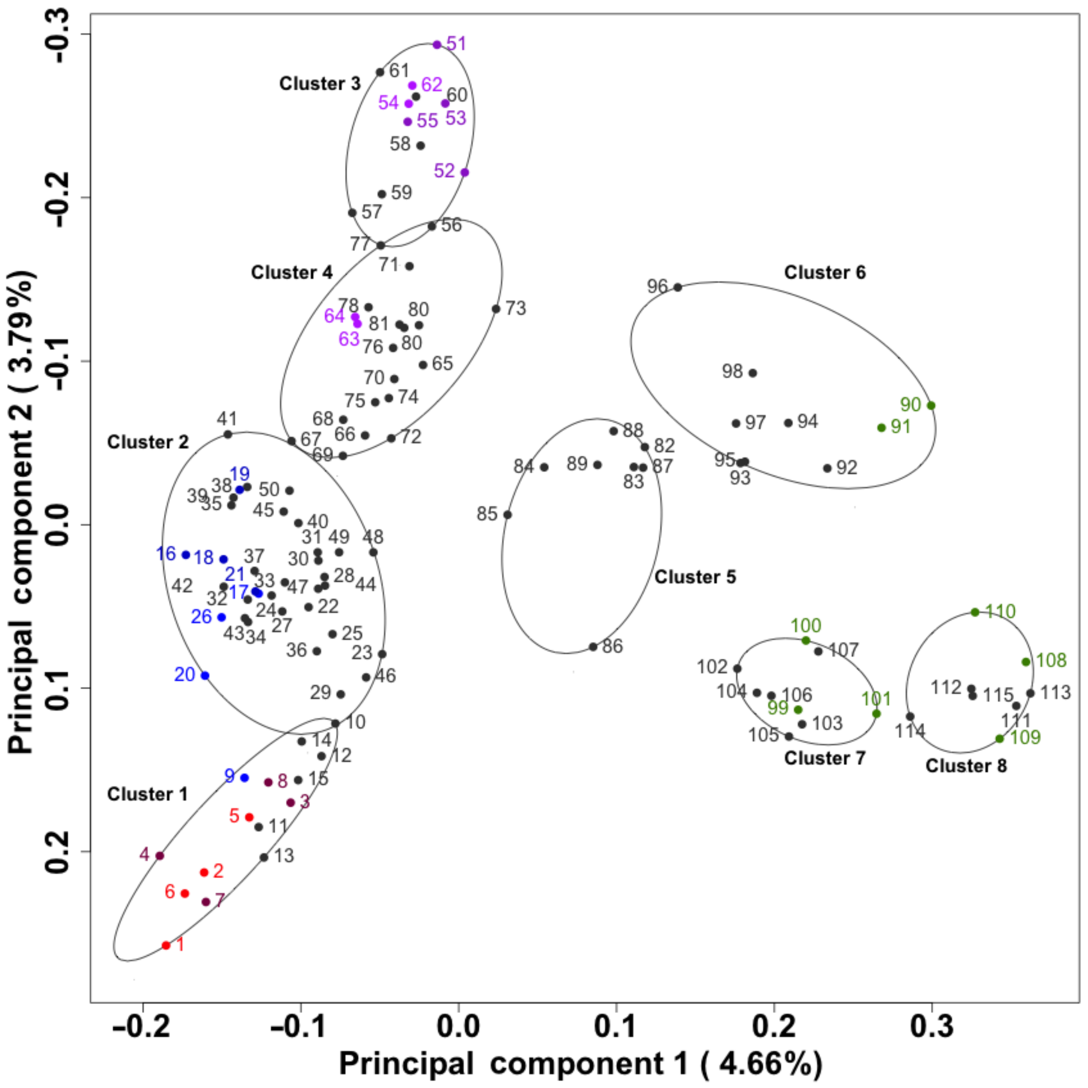


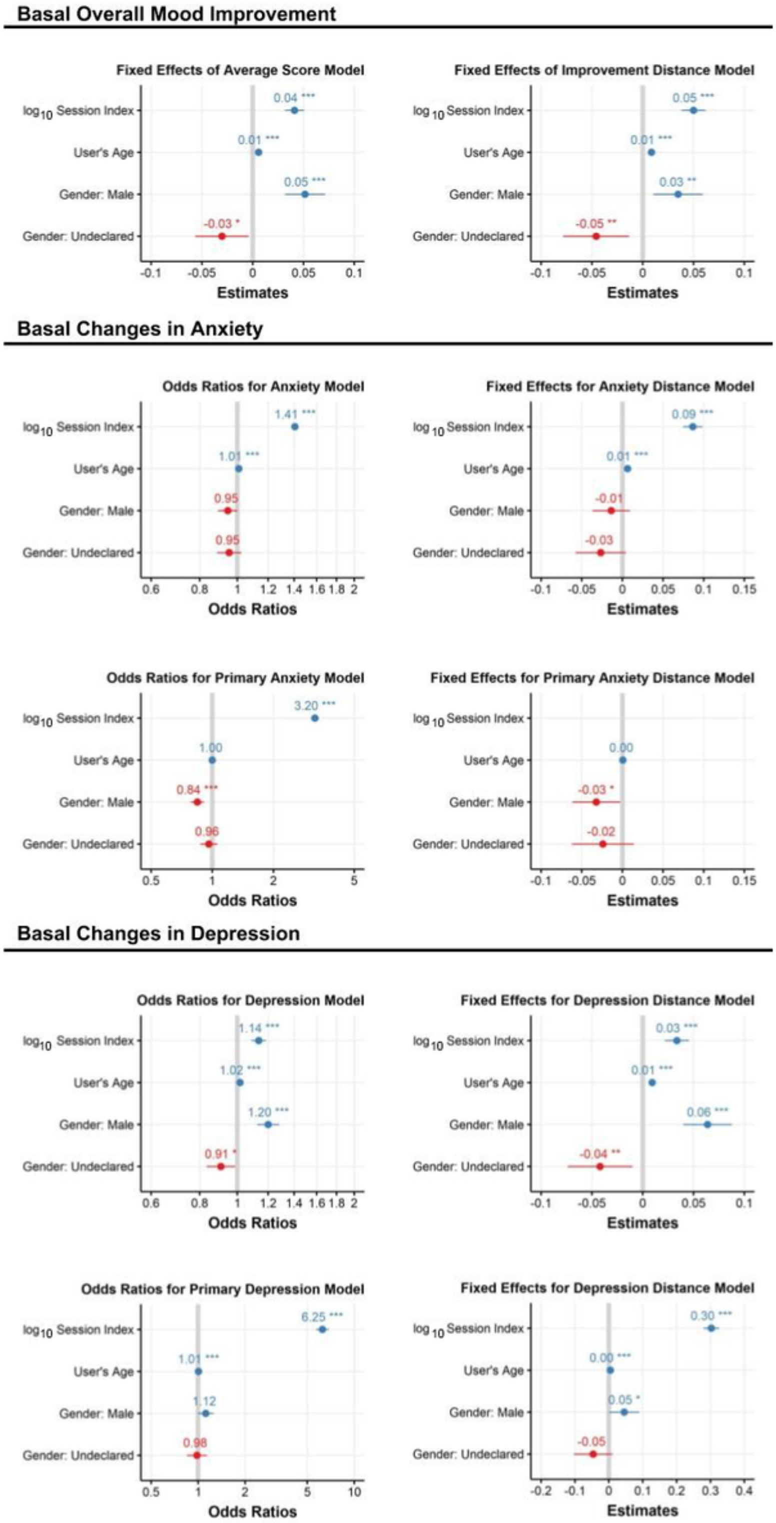
Figure 3. Emotion clustering using both pre- and postactivity emotion endorsements. The points in the plot reflect positions in the first 2 principal components defined by the Bray-Curtis distance between each pre- and postactivity emotional selection. The 8 circular clusters encompassing the emotions were defined by a permutation around medoids analysis technique, in which 8 clusters maximized the average cluster silhouette scores. Cluster boundaries are drawn on the smallest region including all underlying emotions. Emotions are labeled by clinical association such that terms clinically associated with anger are in red and pink, depression in blue, anxiety in purple, and happiness in green.



Mixed-Effects Modeling: Long-Term Use Effect on Preactivity Mood and Emotional State

Using the average preactivity emotional scores, as well as the cluster-based distance measures, as dependent variables, we fit linear-mixed models with session, as well as the important covariates, as independent variables, while accommodating serial correlation emotions. The results using the average preactivity emotional state scores suggest that a statistically significant relationship exists between the number of uses of the app (ie, session index) and the preactivity emotional state, with an elevation in mood (ie, increase in positive emotions) occurring with repeated use of the app. Adjusting for scale, users experience a 2% improvement in mood after their first session, a 4% increase after their 10th session, and a 6% increase after their 100th session. The clinical relevance of this improvement in mood needs to be investigated further. We found that males have an average 2.5% higher (improved) preactivity mood than females and that older users have a more positive mood than younger users. Additional analyses suggested that repeated use of the app resulted in specific improvements in levels of anxiety and depression. After the first 10 sessions with the app—which on average corresponded to a 63.4-day period—users were 82% more likely to report no anxious emotions and 28% more likely to report no depressive emotions. This effect was even more pronounced when we only examined users whose first emotion endorsement reflected anxiety (440%) or depression (1050%). Figure 4 depicts the effect size and statistical significance of the estimated regression coefficients for the analysis models with the average emotional score in the left panels and cluster-based emotion similarity scores in the right panels. The statistical significance (ie, *P* values) were calculated using a Wald-Z statistic approximation. Models fit using a subset of users who reported anxious or depressed emotions in their first session with the app are labeled as *primary* models. The session index is consistently associated with improvements in mood, suggesting, again, that repeated use of the app positively impacts mood.

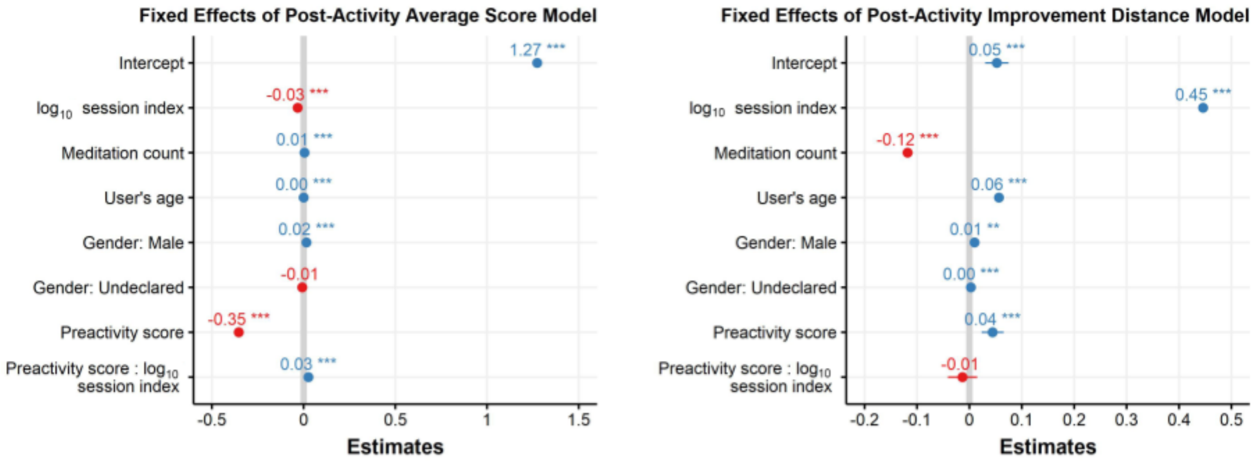
Figure 4. Linear mixed-effects regression coefficient estimates, their SEs, and P values (<.001***, <.01**, and <.05*) for models with the preactivity emotional state as the dependent variable. Analyses with the emotion scoring method as the dependent variable are on the left panels and analyses using distances from clustering as the dependent variable are on the right panels. Generalized Linear Model logit regression models were used with a binary dependent variable indicating if the emotion terms endorsed at a session reflected anxiety (middle panels) or reflected depression (bottom panels).



Mixed-Effects Modeling: Pre- Versus Postactivity Mood or Emotional State

We also fit models that considered the ratio of preactivity to postactivity emotional scores as the dependent variable. Figure 5 plots the regression coefficients resulting from the fits of these models with the ratio of average emotional score pre- to postactivity as the dependent variable (top panel) and the ratio of the distances between the emotions based on the clustering (bottom panel). The results suggest that repeated use of the app leads to increases in improvement of the mood/emotional state achieved through a meditation or mindfulness activity—or rather that the activities seem to lead to larger improvements in mood as the user has more engagements with the app.

Figure 5. Linear mixed-effects regression coefficient estimates, their SEs and P values (<.001***, <.01**, and <.05*) for models with pre- to postactivity change in the emotional state as the dependent variable. An analysis with the standardized change in emotion score pre- to postactivity as the dependent variable is reflected in the top panel, and proximity to the positive emotional clusters as the dependent variable is reflected in the bottom panel.



Discussion

Principal Findings

Our analyses show that repeated engagements with the SBT app are associated with an improvement in users' emotional states over time. In the absence of a randomized control trial, it is difficult to say with certainty that there is a direct causal relationship between the use of the SBT app and emotional state; however, given the large diverse sample size, we believe that the impact of unmeasured covariates on our results (such as external events in the users' lives) is likely to be small, although potential biases in the users of the app may exist. The effect we observed is more pronounced for users who often endorse anxiety or depression when capturing their emotional state at their initial uses. We also found that age and sex covariates are associated with the basal mood or emotional state. Ultimately, our analyses suggest the possibility that guided meditations and mindfulness activities have the potential to be effective ways of reducing anxiety, depression, and stress and ultimately elevating mood, although the ultimate clinical significance of the improvements in the emotional state that we observed needs to be explored. Our analyses did reveal other interesting phenomena. For example, although a minority in our study, males tended to have higher baseline emotional scores and responded better to the SBT app than females. The age of a user was also found to be a significant correlate of the basal emotional state, with older users generally endorsing more positive emotions.

Limitations of the Study

Our analyses are not without limitations, the first and foremost being that there is no control group and comparator app. This makes it difficult to definitively state that guided meditation and mindfulness activities are causally related or responsible for the increase in baseline mood or emotional state over time. However, given the sample size and magnitude of the effect, the significant change in emotional state after immediate and prolonged use of the app suggests that it has potential as an intervention. Another limitation is that all the information we analyzed was self-reported without any oversight by a third party. There could be users who did not follow instructions and entered erroneous emotions to expedite engagement with the meditations. Many of the individuals we did include in our analyses did not record emotions for each and every one of their sessions, resulting in many incomplete observations. Finally, a potential limitation with our analyses is that there could have been a heavy selection bias among the individuals using the app in the sense that they were motivated enough to download it and use it. Thus, this may be an indication that they could be predisposed to responding positively to the app.

Broad Emotional State Transitions

Our use of the emotion clusters and similarity scoring of emotions based on our cluster analyses of those emotions allowed us to explore how often individual users transitioned from one broad set of analogous and almost synonymous emotions to another. On the basis of these analyses, we found evidence that, in general, individual users' emotional states move from negative to positive over repeated uses of the app. We find that anxiety-prone and more depressed individuals benefit from the app more than others. These findings, as with the analyses, need to be verified in more controlled settings, such as randomized control trials, but again suggest that there is promise for the app and related apps in clinical and public health settings.

Future Directions

There are a number of questions that deserve attention beyond those that we addressed with our data. For example, the number of uses of the app may not reflect the total length of time the app was used (eg, a user could engage with the app intensely over a short period of time or stretch their use out over a longer period of time). Assessing the impact of the number of uses versus length of time on outcomes could provide a more detailed insight into the benefits of the app. In addition, it would be good to see if a companion study designed especially for adolescent populations also has a positive effect on their emotions [29]. In addition, special clinical populations may benefit from the app (eg, clinically depressed individuals and individuals with addictions). It would be of value to explore analyses that focus on the impact of large-scale social stressors (eg, school shootings, national election results, and natural disasters) on the use of the app as well as its effects on mood in the wake of stress-inducing events. Geolocation data on users could better define such exposures to social stressors should they be location specific (eg, a natural disaster in a particular state). Finally, as emphasized, it would be ideal to test the utility of the app in bona fide clinical trials to determine which aspects of the app are causally related to improvements in mood and emotional state as well as identifying subgroups of individuals that appear to respond best to particular activities.

As more and more attention is given to the delivery of health care and health maintenance strategies through devices such as smartphones, robots, and telemedicine communications, greater sensitivity to the nuanced effects of these devices should motivate studies of them that are pursued in a comprehensive manner. Such sensitivity and more elaborate studies could also lead to more efficient and sophisticated deployment of these devices and help combat the need for expensive and logistically challenging visits to health care providers.

Conflicts of Interest

Within Stop Breathe & Think, SS and NJS are advisory consultants, JP and JC are cofounders, and JG is an employee. SS, NJS, JP, JC, and JG all hold equity in Stop Breathe & Think.

Multimedia Appendix 1

Assignment and scores for Stop, Breathe & Think selectable emotions.

[XLSX File (Microsoft Excel File), 11KB - mental_v6i5e12617_app1.xlsx]

Multimedia Appendix 2

Histogram of time from first to last recorded session for users with at least ten sessions and six emotional check-ins. On average users participated in sessions with the app over a period of 180 days, with a median use of 119 days, and maximum of 702 days.

[PNG File, 14KB - mental_v6i5e12617_app2.png]

Multimedia Appendix 3

Average user period between sessions. On average a user will interact with the app at least once every 6.34 days, and the majority of users complete at least two sessions per month.

[PNG File, 13KB - mental_v6i5e12617_app3.png]

References

1. GBD 2016 DALYsHALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1260-1344 [FREE Full text] [doi: 10.1016/S0140-6736(17)32130-X] [Medline: 28919118]
2. Asherson P, Kosmas C, Patel C, Doll H, Joseph A. Health-related quality of life and work productivity of adults with ADHD: a UK web-based survey. Eur Psychiatry 2017 Apr 14;41:S352. [doi: 10.1016/j.eurpsy.2017.02.332]
3. Aitken LM, Burmeister E, McKinley S, Alison J, King M, Leslie G, et al. Physical recovery in intensive care unit survivors: a cohort analysis. Am J Crit Care 2014 Dec 31;24(1):33-40. [doi: 10.4037/ajcc2015870]
4. Goetzel RZ, Hawkins K, Ozminkowski RJ, Wang S. The health and productivity cost burden of the "top 10" physical and mental health conditions affecting six large U.S. employers in 1999. J Occup Environ Med 2003 Jan;45(1):5-14. [doi: 10.1097/00043764-200301000-00007] [Medline: 12553174]
5. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014 Nov;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
6. Insel TR. Assessing the economic costs of serious mental illness. Am J Psychiatry 2008 Jun;165(6):663-665. [doi: 10.1176/appi.ajp.2008.08030366] [Medline: 18519528]
7. Smit F, Cuijpers P, Oostenbrink J, Batelaan N, de Graaf R, Beekman A. Costs of nine common mental disorders: implications for curative and preventive psychiatry. J Ment Health Policy Econ 2006 Dec;9(4):193-200. [Medline: 17200596]
8. Remes O, Brayne C, van der Linde R, Lafortune L. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. Brain Behav 2016 Dec;6(7):e00497 [FREE Full text] [doi: 10.1002/brb3.497] [Medline: 27458547]

9. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
10. Iacoviello BM, Murrough JW, Hoch MM, Huryk KM, Collins KA, Cutter GR, et al. A randomized, controlled pilot trial of the Emotional Faces Memory Task: a digital therapeutic for depression. NPJ Digit Med 2018;1 [FREE Full text] [doi: 10.1038/s41746-018-0025-5] [Medline: 30854473]
11. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
12. Lindsay EK, Chin B, Greco CM, Young S, Brown KW, Wright AG, et al. How mindfulness training promotes positive emotions: dismantling acceptance skills training in two randomized controlled trials. J Pers Soc Psychol 2018 Dec;115(6):944-973. [doi: 10.1037/pspa0000134] [Medline: 30550321]
13. Economides M, Martman J, Bell MJ, Sanderson B. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
14. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138. [doi: 10.1037/ocp0000118] [Medline: 29723001]
15. Coulon SM, Monroe CM, West DS. A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
16. Waltz E. Pear approval signals FDA readiness for digital treatments. Nat Biotechnol 2018 Dec 6;36(6):481-482. [doi: 10.1038/nbt0618-481] [Medline: 29874220]
17. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar 10;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
18. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. J Ment Health 2011 Dec;20(6):509-524. [doi: 10.3109/09638237.2011.608746] [Medline: 21988230]
19. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
20. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and impact of real-world clinical data for the practicing clinician. Adv Ther 2018 Nov;35(11):1763-1774 [FREE Full text] [doi: 10.1007/s12325-018-0805-y] [Medline: 30357570]
21. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler H, Goettsch W, et al. Good practices for real - world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR - ISPE Special Task Force on real-world evidence in health care decision making. Value Health 2017 Dec;20(8):1003-1008 [FREE Full text] [doi: 10.1016/j.jval.2017.08.3019] [Medline: 28964430]
22. Monti S, Grosso V, Todoerti M, Caporali R. Randomized controlled trials and real-world data: differences and similarities to untangle literature data. Rheumatology (Oxford) 2018 Dec 1;57(57 Suppl 7):vii54-vii58. [doi: 10.1093/rheumatology/key109] [Medline: 30289534]
23. Baer R. Mindfulness training as a clinical intervention: a conceptual and empirical review. Clin Psychol 2003;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
24. Zoogman S, Goldberg S, Hoyt W, Miller L. Mindful Well-Being. 2015. Mindfulness interventions with youth: A meta-analysis URL: http://www.mindful-well-being.com/wp-content/uploads/2014/07/Zoogman-et-al-2014-meta-anlysis.pdf [accessed 2019-04-11] [WebCite Cache ID 77YDPwxOO]
25. Zoogman S, Goldberg SB, Hoyt WT, Miller L. Institut de recherche pour le développement. 2014 Jan 15. Mindfulness Interventions with Youth: A Meta-Analysis URL: http://hal.ird.fr/ird-01887318/document [accessed 2019-04-11] [WebCite Cache ID 77YE7TtR6]
26. Dixon, Philip. VEGAN, a package of R functions for community ecology. J Veg Sci 2003;14(6):927-930. [doi: 10.1111/j.1654-1103.2003.tb02228.x]
27. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
28. Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. J Math Model Algor 2006 Mar 3;5(4):475-504. [doi: 10.1007/s10852-005-9022-1]
29. Monto M, McRee N, Deryck F. Nonsuicidal self-injury among a representative sample of US adolescents, 2015. Am J Public Health 2018 Aug;108(8):1042-1048. [doi: 10.2105/AJPH.2018.304470] [Medline: 29927642]

Abbreviations

FDA: Food and Drug Administration
GLM: Generalized Linear Model

LME: Linear Mixed-Effects
PAM: Partitioning Around Medoids
PCoA: Principal Coordinates Analysis
SBT: Stop, Breathe & Think

Edited by J Torous; submitted 26.10.18; peer-reviewed by R Vanderkruik, U Subnis; comments to author 11.12.18; revised version received 02.02.19; accepted 24.03.19; published 08.05.19.

Please cite as:
Athanas AJ, McCorrison JM, Smalley S, Price J, Grady J, Wehner P, Campistron J, Schork NJ
Association Between Improvement in Baseline Mood and Long-Term Use of a Mindfulness and Meditation App: Observational Study
JMIR Ment Health 2019;6(5):e12617
URL: https://mental.jmir.org/2019/5/e12617/
doi:10.2196/12617
PMID:31066704

Original Paper

Gaming With Stigma: Analysis of Messages About Mental Illnesses in Video Games

Manuela Ferrari[1*], MHSc, PhD; Sarah V McIlwaine[1,2*], BA; Gerald Jordan[3], PhD; Jai L Shah[1,2], MD; Shalini Lal[1,4,5], PhD, OTR; Srividya N Iyer[1,2], PhD

[1]Prevention and Early Intervention Program for Psychosis, Douglas Mental Health University Institute, Montreal, QC, Canada
[2]Department of Psychiatry, McGill University, Montreal, QC, Canada
[3]Yale Program for Recovery and Community Health, Yale University, New Haven, CT, United States
[4]Carrefour de l'innovation et de l'évaluation en santé, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada
[5]School of Rehabilitation, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada
[*]these authors contributed equally

Corresponding Author:
Manuela Ferrari, MHSc, PhD
Prevention and Early Intervention Program for Psychosis
Douglas Mental Health University Institute
Pavillon Burland
6875 Blvd LaSalle
Montreal, QC, H4H 1R3
Canada
Phone: 1 5147616131 ext 2087
Email: manuela.ferrari@douglas.mcgill.ca

Abstract

Background: Video game playing is a daily activity for many youths that replaces other media forms (eg, television); it serves as an important source of knowledge and can potentially impact their attitudes and behaviors. Researchers are, thus, concerned with the impact of video gaming on youth (eg, for promoting prosocial or antisocial behavior). Studies have also begun to explore players' experience of gameplay and video game messages about violence, sexism, and racism; however, little is known about the impact of commercial video games in the sharing and shaping of knowledge, and messages about mental illness.

Objective: The aim of this review was to identify how mental illness, especially psychosis, is portrayed in commercial video games.

Methods: We performed keyword searches on games made available between January 2016 and June 2017 on Steam (a popular personal computer gaming platform). A total of 789 games were identified and reviewed to assess whether their game content was related to mental illness. At the end of the screening phase, a total of 100 games were retained.

Results: We used a game elements framework (characters, game environment/atmosphere, goals, etc) to describe and unpack messages about mental health and illness in video games. The majority of the games we reviewed (97%, 97/100) portrayed mental illness in negative, misleading, and problematic ways (associating it with violence, fear, insanity, hopelessness, etc). Furthermore, some games portrayed mental illness as manifestations or consequences of supernatural phenomena or paranormal experiences. Mental illness was associated with mystery, the unpredictable, and as an obscure illness; its treatment was also associated with uncertainties, as game characters with mental illness had to undergo experimental treatment to get better. Unfortunately, little or no hope for recovery was present in the identified video games, where mental illness was often presented as an ongoing struggle and an endless battle with the mind and oneself.

Conclusions: The game elements of the identified commercial video games included mental illness, about which many perpetuated well-known stereotypes and prejudices. We discuss the key findings in relation to current evidence on the impact of media portrayals of mental illness and stigma. Furthermore, we reflect on the ability of serious video games to promote alternative messages about mental illness and clinical practices. Future research is needed to investigate the impact that such messages have on players and to explore the role that video games can play in fostering alternative messages to reduce the stigma associated with mental illness.

(JMIR Ment Health 2019;6(5):e12418) doi:10.2196/12418

KEYWORDS

mental disorders; social stigma; video games

Introduction

Background

More than 1.8 billion people play video games worldwide [1]. In a US national survey, 1102 teens aged 12 to 17 years described video gaming as a daily activity [2]. Contrary to a popular belief that views video game playing as a solitary activity, most youth (76%) played video games with others in person or on the Web, and only 24% played alone. In addition to console games (Xbox, PlayStation, or Wii), youth reported playing games on their desktop or laptop computers. First-person shooters, action, and sport games were the top types of video games played; however, most youth did not limit themselves to just one game genre [2].

Given the popularity of video games among young people, researchers have been studying the role that this technology plays in promoting prosocial and antisocial behavior among players. The vast majority of research has focused on the negative impact of video games, describing their potential to lead to aggression, addiction, and depression [3,4]. However, new emerging research challenges this body of work. Notably, Zendle critically analyzed the relationship between violent content in video games and aggressive behaviors [5]. Zendle found that it was players' frustration with game complexity and difficulty that led them to behave more aggressively after playing rather than the violent content within games.

Emerging literature has also identified the benefits of playing video games in promoting better attention, memory and problem-solving skills, and enhancing one's ability to cope with failures, manage emotions, and socialize [6]. Furthermore, novel research has focused on the impact of commercial games as health interventions, psychotherapy tools for assessment, and in teaching social skills [7]. In addition, a few reviews have been published to evaluate the effectiveness of video game technology in the treatment of different mental health conditions (eg, depression and anxiety) [8-10].

Studies have also begun to explore players' experience of gameplay and video game messages about violence, sexism, and racism [11-14]. However, little attention has been devoted to messages about the experience of mental illness in commercial video games [15-17], the impact they may have on attitudes toward and understandings of mental illness, and how they could perpetuate stereotypes and/or stigma about mental illness. Shapiro and Rotter [16] reviewed 96 popular commercial video game characters (highest-selling video games from 2011 to 2013) comparing them with mental illness stereotypes identified in movies [16,18]. Almost all identified video games (93 of 96) depicted at least one character with mental illness while linking mental illness to dangerous and violent behaviors.

Apart from a few published works that looked at mental illness messages in video games [15-17], most of our current knowledge about media portrayals about mental illness concern movies, television (TV), and cartoons. Since the 1990s, researchers have studied the impact of media (eg, TV, newspapers, and cartoons) portrayals of mental illness on the public [19-22] and most recently on children [23]. Specifically, these depictions perpetuated the stereotype that people suffering emotionally might be *mentally ill*, dangerous, or violent [18,24,25]. The portrayal of persons with mental illness as dangerous, incurable, and incapable of controlling their illness has been shown to promote negative attitudes about mental illness and problematic behaviors toward persons with mental illness [26]. In their study on verbalizations (labeling) about mental illness in Disney animated films, Lawson and Fouts [23] found that of the 40 full-length animated feature films produced between 1937 and 2001, 34 films depicted mental illness using terms such as *crazy* or *nuts*. Specifically, these terms were used to marginalize and denigrate the characters to whom they referred. Under the influence of media messages, children and youth can begin stigmatizing mental illnesses from a relatively early age.

As described, negative portrayals of mental illness in traditional media are a matter of social concern; however, the impact of mental illness–related content in video games is unknown. To unpack how reality is presented via video games and the potential that these presentations have for shaping youth perceptions, it is necessary to understand the *narrative rhetoric* of game content [27], and the messages that games portray.

Review Aim

The aim of this exploratory study was to examine the labels used and overall messages about mental illness, especially psychosis, portrayed in video games currently on the market. We explored not only how the experience of mental illness was portrayed in these games but also how treatment and the settings of care were described and represented. We focused on psychosis because it is among the most stigmatized mental illnesses in TV shows and movies, mainly because its unique manifestations through hallucinations and delusions capture the imagination of writers, producers, and audiences [28,29]. Our specific research question was as follows: *How is mental illness, especially psychosis, as well as its context (eg, treatment and settings of care) portrayed in commercial video games?*

Methods

Epistemic Framework: Discourses, Messages, and Mental Health Stigma

Informed by critical discourse analysis [30], this work explores the language used to describe and explain mental illness in recently released commercial video games. Critical discourse analysis examines the form, structure, and content of discourse, including the wording chosen and used in a text, to unpack direct and indirect messages sent to the audience.

Theoretically, stigma about mental illness is generated and shaped by labeling and stereotyping. Labeling refers to the use of a term to define a person that impacts people's attitudes, beliefs about, and behaviors toward that person [31]. Labels ascribed to a person with a mental illness can generate positive or negative connotations about that individual. For example, describing a person with psychosis using words such as *dangerous* can denigrate the person and the person's illness experience. Stereotyping is the application of a belief about a person that can shape one's behavior toward that person. Stereotypes can arise from labeling, cultural beliefs, and direct or indirect exposure to a person or a group. If, for example, a person with a history of mental illness commits a crime, especially a violent one, media reports may use terms such as *psychotic* or *out of control* to describe the perpetrator. In such a scenario, a connection is made among the person in question, the person's mental illness, the labels used, and the crime committed, which perpetuates the stereotype that people with mental illness are violent and dangerous. A review of epidemiological studies shows that the majority of people with mental disorders do not engage in violence against others and that they are actually more at risk to harm themselves than others [32].

Mental illness stigma can lead to 2 problematic behaviors: marginalization, or the separation between *us* and *them* [33], and discrimination, or behaviors that deny social participation or access to resources. Research has shown that stereotypical beliefs about mental illnesses and persons with mental illnesses contribute to social discrimination, including in the domains of jobs and housing [34-36]. Furthermore, it has been suggested that adults' stereotypical beliefs about mental illness may have originally been acquired through media exposure in childhood and teen years [37-39].

Search Strategy

On Steam, a popular personal computer (PC) gaming platform, we performed keyword searches using the following terms: "asylum," "insane," "crazy," "mad," "madness," "mental," "psycho," "psychotic," "psychosis," and "schizophrenia" on games released between January 2016 and June 2017, to capture 1.5 years of commercial video game production and dissemination. We selected the Steam platform because it has access to a wide variety of gaming genres and indie games, with PC game sales on this platform being almost as high as all console game sales [40]. In addition to popular and well-known terms or labels associated with psychosis such as "crazy" and "insane," we searched for formal/medical terms such as "schizophrenia" and "psychosis." A similar key-terms strategy, a mix of clinical terms and common mental illness labels, was also used in the study by Shapiro and Rotter [16].

A total of 2 reviewers (MF and SM) systematically used the selected keywords on Steam to identify games for inclusion in the analysis. Games were included for full-text review if (1) they were published on Steam between January 2016 and June 2017 and (2) regardless of genre, they featured and/or presented mental illness and/or mental health content. MF and SM first performed a preliminary screen to identify which games were available within the aforementioned period, which yielded a total of 789 games. Furthermore, the reviewers systematically reviewed the Steam game descriptions to exclude games in whose synopses the identified keywords appeared, but not in the context of mental illness (eg, only in the title as in the game *Crazy Fishing*). If a game did meet the criteria, its game information was extracted from Steam's website and imported into an electronic database for analysis. On the basis of this search criteria, 100 games were retained for data analysis.

Data Extraction and Coding

We created an electronic database of all retained games. Metadata such as game title, year, genre, and synopsis were entered in an Excel file. The Excel file was then exported into Atlas.ti version 8 (Atlas) [41] to perform thematic analysis. We followed Braun and Clarke's [42] steps for thematic analysis, informed by an inductive process that began by generating initial codes and themes, and then creating a coding manual. According to the authors [42], thematic analysis is a flexible analytic method that can be effectively applied across a range of epistemological approaches, including discourse analysis [43]. Codes and themes were reviewed and redefined by exploring relationships between code-level and theme-level preliminary analysis. We interpreted the meaning of each identified theme in relation to our research question: *How is mental illness, especially psychosis, as well as its context (eg, treatment and settings of care) portrayed in commercial video games?* Finally, we further unpacked and finalized the analysis by comparing the identified themes with game elements such as genre, game characters, story, atmosphere, game goals, and the game environment, and we used this framework to report our findings.

Results

Video Game Taxonomy and General Overview

Information about the 100 games retained for analysis is presented in Table 1 and Figure 1. Overall, the keywords that produced the most hits were "insane" (198/789), "crazy" (244/789), and "mad"/"madness" (244/789), which may reflect the colloquial use of these words to describe excitement, fun, or a "wild" experience. For instance, "crazy" was mainly associated with puzzle games for children and younger audiences. Words such as "crazy," "insane," and "madness" were used to convey characters who do not abide by rules or environments that are chaotic or creative. For example, *Crazy Saloon VR* (Monsieur K, 2016) is described as follows:

> *Discover the Old West with this quirky saloon simulator...There is a free mode where you'll be able to roam freely around the saloon shooting things and stuff, fully express your craziness creativity!*

The game *TRATEL64* (UWILMOD, 2017) explicitly equated "insane" with "eccentric":

> *Slice, Snipe, Stab, and Shoot your way through 64 insane targets! Experience the insane (or eccentric if you prefer) personalities of a truly twisted world!*

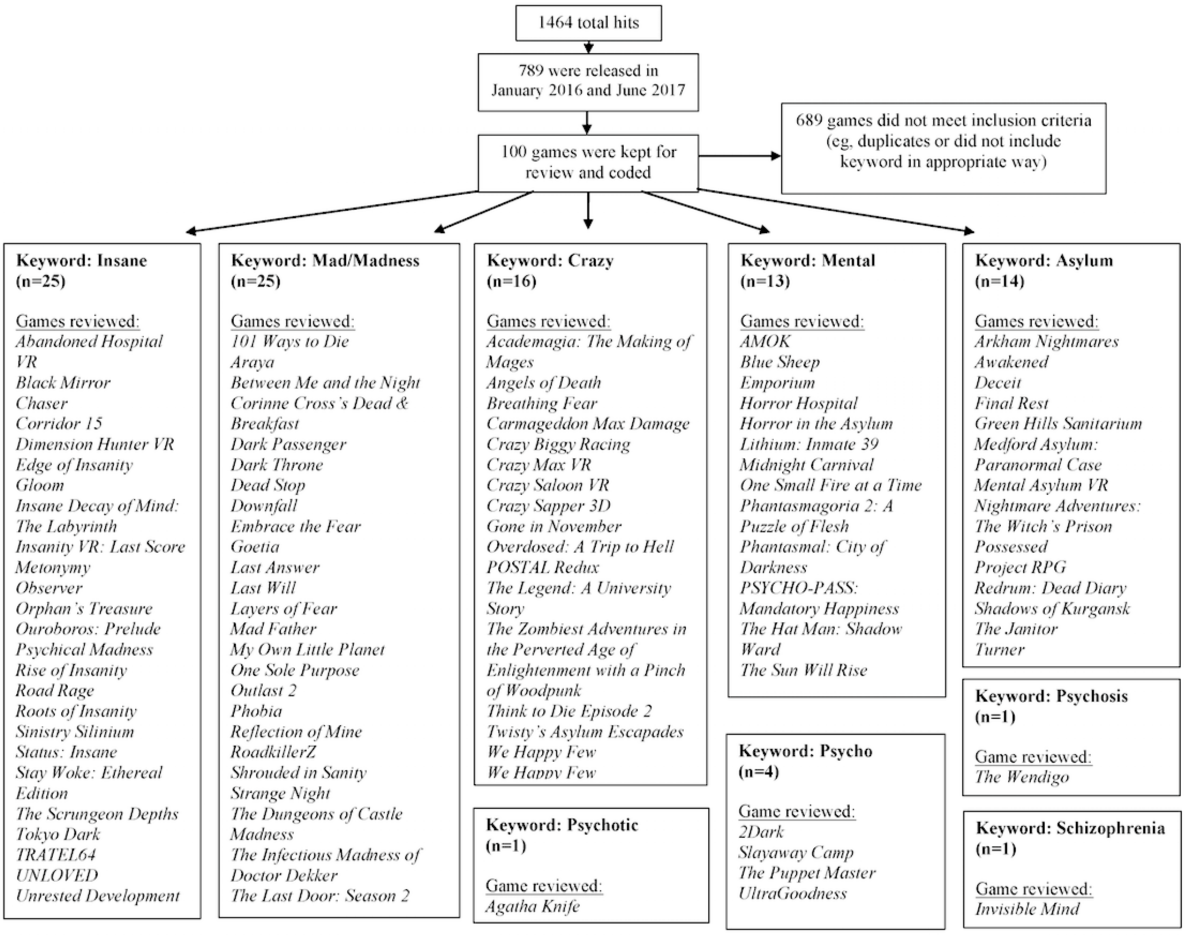
Table 1. Game analytics summary for 100 games.

Keyword[a]	Games per keyword, n	Total number of gamer reviews posted on steam		Average cost of games ($CAD), mean	Games with viewing restrictions (age restriction or "not safe for work"), n (%)
		N	Median		
Insane	25	4221	44	9.81	11 (44)
Mad/Madness	25	18,702	78	10.01	5 (20)
Crazy	16	7139	80	12.38	4 (25)
Asylum	14	19,088	20.5	5.42	7 (50)
Mental	13	3704	58	8.84	3 (23)
Psycho	4	621	89	10.29	0 (0)
Psychosis	1	16	—[b]	5.49	1 (100)
Psychotic	1	88	—	12.99	0 (0)
Schizophrenia	1	23	—	0.60	0 (0)

[a] Information accurate as of September 2018 (see Multimedia Appendix 1 for more details).

[b] Not applicable.

Figure 1. Video games extraction summary.



Similarly, the description of *Crazy Sapper 3D* (Aratog LLC, 2016) noted the following:

> *A mad retired General named Boris has hidden a devastating acid vaccine on a secret base...There's only one man standing in the way of Boris' dream of creating a super-bomb and enslaving the entire world: a crazy sapper named Max!*

In this example, Boris, the antagonist, and Max, the protagonist, both embody being "mad" and "crazy"—stereotypical terms associated with mental illness. The frequency of their occurrence and their use as selling points in the game synopses suggest that terms such as “crazy” and “mad” were accorded (by game creators) or had come to acquire (among gamers) some desirable connotations.

On the other hand, keywords that were more medical in nature produced fewer hits, with “psychotic” resulting in only 4 hits out of 789, “psychosis” in only 3 hits, and “schizophrenia” in only 1 hit. Unlike the word “crazy,” medical terms such as “psychosis” and “schizophrenia” can bring a level of formality and seriousness to games that gamers may not seek, leading to their avoidance or exclusion by game marketers [44,45].

Although the style of game play varied, the genres of games were relatively homogeneous, with most games pertaining to action and adventure genres (83%, 83/100), and a majority (68%, 68/100) being tagged by users as containing horror, violence, and gore.

Mental Illness and Game Elements

Using the game elements framework [46], we unpacked and described messages about mental illness presented in the identified games. Specifically, we explored these messages based on the following structure: themes associated with the main and secondary characters of the game, the game atmosphere (including location, time period, and overall environment), game goals, and how the game described mental illness and its experience. For each section, we provided an estimate of the percent of games that pertain to the themes presented. This estimate accounts for the potential overlap among themes.

Mental Illness and Game Characters: Being Violent, Lost, Lonely, or Helpless

Of the 100 games we reviewed, almost all game characters (whether the main character or secondary characters) were described as “psychotic,” “crazy,” “mad,” “eccentric,” “paranoid,” “unusual,” “evil,” and “insane.” In approximately 30% to 35% of games, characters were portrayed as either *being violent* and/or as *being lost, lonely, or helpless*.

In the reviewed synopses, in approximately 16% of games, the characters were described as *violent or aggressive*. This was especially true of characters who were portrayed as having psychosis, whether explicitly or implicitly, through behaviors such as hearing voices or seeing visions. For instance, violence and “being psychotic” were paired and emphasized in *UltraGoodness* (Rasul Mono, 2017; 80% of 89 gamer reviews were positive), a game that requires players to fight “a fricking army of psychos” and “splatter blood all over the levels.” Although apparently comedic and satirical, the game creates a connection between violence and persons with mental illness (who, in this case, are victims). A similar connection was observed in *Slayaway Camp* (Blue Wizard Digital, 2016; 96% of 409 gamer reviews were positive), whose description reads:

> *Slayaway Camp is a killer puzzle where you control Skullface, a psychotic slasher bent on slaughtering camp counsellors...this adorable murderer slides around isometric puzzle levels decapitating, squashing, and perforating his bloody victims.*

In approximately 15% of games’ synopses, protagonists were described as *lost, lonely, or helpless* in some way. Often, protagonists had to find characters who were lost or were themselves lost. In some games, the journey of mental illness and recovery was therefore represented through a character finding what was lost and overcoming challenges that drive them to “insanity.” For instance, players of *My Own Little Planet* (Lucas Parise, 2017; 63% of 11 gamer reviews were positive) must help a “lost boy who is trapped on a dark and strange world.” They have to “fight [their] deepest fears” to “put an end to [the] infinite loop of madness and fright.”

This theme of finding what was lost was also emphasized in *Tokyo Dark* (Cherrymochi, 2017; 91% of 373 gamer reviews were positive), wherein the protagonist, Detective Ito, in her quest to find her missing partner, begins to “question her own sanity.”

Mental Illness and Atmosphere

In unpacking relationships between depictions of mental illness and game atmosphere, approximately 75% to 80% of games had game settings and ambiances that pertained to themes of mental illness.

Game Setting

Overall, it was interesting to notice how some games (10%) depicted time periods reminiscent of the Victorian and Edwardian eras, referring to mental health care in historic terms (eg, describing mental health institutions as asylums or sanitoriums). For instance, *Arkham Nightmares* (Tim Rachor, 2016; 20% of 10 gamer reviews were positive) takes place in the “Lovecraftian Arkham Sanitorium” and *Layers of Fear* (Blooper Team SA, 2016; 91% of 6831 gamer reviews were positive) takes place in a “Victorian-era mansion.”

The troubling and ethically questionable history of the treatment of mental illness provided a suitably macabre context for horror games such as *The Hat Man: Shadow Ward* (Game Mechanics, 2016; 67% of 2937 gamer reviews were positive), which tells “the real-life story of the events that took place at the Canton State Insane Asylum fifty years ago, as reported by those who survived.” Its players are “thrust into a living nightmare” in which “electricity and communication with the outside world has been cut off, and the asylum is being terrorized by supernatural beings.” Obviously, supernatural beings did not roam the halls of the Canton State Insane Asylum, but the asylum did exist and had a problematic history. Using it as a setting served to enhance the believability of the game’s story.

Furthermore, mental health care in futuristic games (6%) also described the health care setting in negative ways. In these dystopian futures, the “mentally ill” are restrained, contained, and constantly surveilled. In one such game, *Observer* (Blooper Team SA, 2017; 82% of 1704 gamer reviews were positive), players assume the role of an Observer, “the new front line of neural police” and must “hack into the jagged minds of the insane.” The synopsis further describes the game’s setting as follows:

> *In this future, anything you think, feel, or remember can be used against you in a court of law...The year*

> *is 2084. Those who live have turned to drugs, VR, neural implants- anything to distract themselves from this new reality.*

Therefore, in both past and future scenarios (16% of games), individuals with mental illness are depicted as dangerous, in need of containment and separation from society, and requiring treatment that involves constant surveillance by doctors or authority figures, usually in asylums.

Some games (19%) used hospitals and asylums as elements of the environment and story. Asylums and hospitals were described as “abandoned,” “decrepit,” “dark,” or in a “state of disarray.” Asylums were also associated with potentially illegal or strange activity. For example, in *Janitor* (VOS Gaming, 2016; 20% of 25 gamer reviews were positive), the player, Max, is a janitor “sent to an asylum by his boss to repair a broken pipe and a broken electric panel.” The synopsis states the following:

> *Max knows the asylum is abandoned...His boss warned him not stay long there. Once arrived he quickly notices that something is wrong. Blood everywhere and strange noises can be heard. Short time after looking around a went-wrong experiment of the government rushes to him.*

In such games, environments (ie, characters’ homes, cities, or minds) were often described as unsettling, disturbing, and in some form of chaos. For instance, *Insanity VR: Last Score* (Threevol, 2017; 86% of 103 gamer reviews were positive) is set inside the mind of a psychiatric patient:

> *The game tells the story of a grey future where...it is possible to enter another’s subconsciousness and treat their psychological problems...The player enters the subconscious mind of a former instrument maker. Shortly after the entry...he is locked inside. During the search for a way out, the terrible past of the instrument maker is gradually revealed and the player discovers that he is not alone.*

Contrastingly, the game *Horror Hospital* (KNP, 2016; 26% of 156 gamer reviews were positive) takes place in the protagonist’s mind, but the story included the “patient’s recovery and awakening.” In addition, the game setting was also associated with dark, disorganized, supernatural, or disturbing places.

In addition to abandoned, disturbing, and broken spaces, 27% of games were set in dangerous, unpredictable, and violent cities or worlds on the verge of collapse. Houses were portrayed as scary and dark and described as “littered with corpses and other horrors” or “teeming with infernal creatures.” Cities were portrayed as dangerous, “ravaged by war,” “drug-fueled,” and “perilous.” For instance, *Shrouded Sanity* (Steve Gal, 2016; 85% of 221 gamer reviews were positive) is set in an estate filled with “violent inhabitants”:

> *A strange and repelling madness took hold of the servants...and they will attack on sight...Discover the secrets of the estate, as you look for answers among the violent inhabitants.*

A similarly violent environment is depicted in *POSTAL Redux* (Running With Scissors, 2016; 90% of 944 gamer reviews were positive):

> *Crazed gunmen out for your blood await you around every corner. The only choice is clear: Get them before they get you. Fight back with a devastating arsenal as you make your way through a violence-stricken town.*

Game Ambiance

Game ambiance refers to the overall “feel” and atmosphere that a video game portrays.

Game synopses (17%) boasted creepy and eerie music and emphasized the supernatural. One game described an asylum as being “full of paranormal activity.” Paranormal activity was a salient and recurring theme, with 11% of games describing buildings or hospitals as being cursed or haunted by ghosts and “otherworldly horrors.” Game environments were described using terms such as “living nightmare” or as spaces where “fear and horror intertwine.”

Creepy and horror-filled atmospheric elements were particularly striking in *Layers of Fear*, a game that requires players to take on the role of an “insane” painter who must navigate a dark, ever-changing mansion to uncover clues about his wife’s death:

> *Delve deep into the mind of an insane painter and discover the secret of his madness...Uncover the visions, fears and horrors that entwine the painter...A sense of insanity means each turn of the camera may completely change the look of your surroundings.*

Creepy imagery was also a key feature of *Medford Asylum: Paranormal Case* (Mzone, 2016; 46% of 43 gamer reviews were positive), where its protagonist “must find what forced workers to stop the renovation. They are frightened by this ‘haunted’ place.”

Horrific, foreboding descriptors were used not only for games’ environments but also for the mental illness itself. For instance, *The Infectious Madness of Doctor Dekker* (D’Avekki Studios Ltd, 2017; 81% of 191 gamer reviews were positive) requires players to explore the “shadow reality” that their “patients inhabit.”

Mental Illness and Game Goals

Video game goals refer to the objectives that players must achieve to progress, complete, or win the game. In exploring the relationship between mental illness and game goals, we identified the following themes: solving/uncovering a mystery, survival, finding an escape, and restoring normalcy in approximately 65% to 70% of games.

A large portion of games (30%) dealt with solving puzzles or resolving mysteries. In addition to deploying puzzle solving as a game mechanism, many such games required players to solve murder cases, discover suspects, or piece together “mysterious circumstances.” For instance, the objective of *Rise of Insanity* (Red Limb Studio, 2017; 80% of 287 gamer reviews were positive) is as follows:

> *As a doctor of psychology, you will discover the gloomiest parts of the human brain. Overcome your fear and find out what really happened to your family...No one knows what has happened to your family...The only suspect is your patient on whom you are testing your experimental treatment methods. Who is responsible for everything that has happened?*

The goals of other games involved understanding or discovering phenomena or scenarios that were initially unclear to the games' characters or players. Examples included game goals such as "discovering the secret to his madness," "understand what is going on in his mind," or "discover who is real." A case in point is *Reflection of Mine* (Redblack Spade, 2017; 85% of 89 gamer reviews were positive) that takes place in the mind of a young girl with a dissociative identity disorder. Its synopsis reads:

> *The entire game takes place in the broken mind of Lilly Witchgan, and the goal is to discover who is real—Lilly herself or one of her many personalities...Here, madness is represented as it really is—Lilly is a foreigner inher own body...*

Other games involved finding what is lost to the character. In one such game, players are tasked with finding the character's missing client. In another, the protagonist has to search for a friend or family member. However, most games whose plots involved a mysterious element dealt with uncovering or discovering forgotten parts of a character's past or a traumatic event.

Survival was a significant theme in the games reviewed. Many games (14%) required players to survive passage through dangerous areas, the onslaught of enemies, or even, in the case of one game, "the horrors within." *Abandoned Hospital VR* (Munsunghyun, 2016; 67% of 3 gamer reviews were positive) requires players to find objects to "stay alive and to escape from the hospital." Overcoming concrete challenges, obstacles, and traps or some more abstract conceptual tests were also common tropes. In *Between Me and the Night* (RainDance LX, 2016; 68% of 81 gamer reviews were positive), players have to solve puzzles to reveal new places and "overcome the shadows of the night."

Game goals often (14% of games) involved escaping, either symbolically (eg, "finding an exit to this madness") or concretely (eg, "escape the asylum"). For instance, in *Awakened* (Jesper Michael Petersen, 2016; 0% of 3 gamer reviews were positive), escaping the "asylum for the criminally insane" involves "escaping the voices in [the protagonist's] head." In many games, the escape to avoid being affected was from a hospital or asylum, with some requiring players to flee the "lairs of psychopaths" or the "clutches of an evil doctor." In *Status: Insane* (Frostbullet, 2017; 95% of 20 gamer reviews were positive), escaping an asylum is the sole objective and story line:

> *Igor must escape from a mental asylum but it's not going to be easy. There are two obvious reasons: Igor is constrained by a straitjacket; The asylum's corridors are filled with armed guards (and other obstacles). In addition to the dangers coming from outside, Igor is going through a war inside his head: His Imaginary Friend is constantly demanding him to cancel his silly escaping attempt.*

This game robustly outlines many mental illness stereotypes—that escaping mental illness is futile, that mental health institutions are dangerous, and that patients in these institutions are constrained and confined by force.

Finally, the retention or regaining of sanity was an important goal in the story lines of some games (10%), describing the need to "end the infinite loops of madness and fright" or "maintaining mental health while facing the worst." For instance, the synopsis of *Phantasmal: City of Darkness* (Eyemobl, 2016; 41% of 115 gamer reviews were positive) stated:

> *You are fragile. Not just physically, but mentally as well. The grotesque creatures will challenge your very sanity and losing your mind can be a fate worse than death!*

Here, the maintenance of sanity is deemed difficult and insanity is deemed the worst consequence. The fragility of sanity was also referenced in *Between Me and the Night*, which boasted an experience that walked players through the "thin path between sanity and madness."

Recovery and normalcy too were described as fragile states in *Phantasmagoria 2: A Puzzle of Flesh* (Sierra, 2016; 80% of 26 gamer reviews were positive):

> *[The protagonist] has a steady job...a lovely girlfriend. He's been out of the mental hospital for exactly one year. All [the protagonist] wants is to live a normal, happy life, but something seems to have other plans...Strange events, inexplicable and terrifying, begin to happen all around him. [The protagonist] begins to doubt his own sanity, and the very fabric of reality.*

Video games, thus, commonly depicted mental health as fragile and unpredictable and recovery as difficult to achieve and maintain.

The Lived Experience of Mental Illness as a Central Element of the Game

We explored whether and how mental illness was used as a central element in games. Only a minority of games' descriptions (12%) included more medical terms such as "dissociative identity disorder," "delusions," "depression," and "anxiety." However, closer examination revealed that even when these terms were used, diagnostic labels seemed to be misapplied or used without a full understanding of their medical meanings. This was especially true for terms associated with psychosis and dissociation. Psychosis and other severe mental illnesses were typically associated with paranoia, lack of control, being followed by a "dark presence," having an infection, being a "foreigner in her own body," having a "fragmented sense of self," being "locked in his own mind," being "haunted by nightmares," and/or supernatural and paranormal experiences. For instance, in *The Wendigo* (Warka, 2017; 93% of 16 gamer reviews were positive), players have to participate in a paranormal investigation to uncover the mystery behind an evil spirit that caused psychosis. The game characterizes psychosis as being expressed through violent and unpredictable behavior and the "insatiable desire to eat human flesh." The association between psychosis and the supernatural is also salient in *Redrum: Dead Diary* (Anarchy Enterprises, 2016; 55% of 27 gamer reviews were positive):

> *Rose sees dead people who have been killed unjustly, with their mortal wishes left unfulfilled. Her super-practical father refuses to believe her ghostly visions and puts her in an asylum, where she falls into the clutches of the evil Dr. Sigmund Fraud. Help Rose use her psychic powers to solve gruesome murders and outwit a homicidal maniac in this spine-chilling mystery.*

Experiences of depression and anxiety were associated with suffering, darkness, and being at the "bottom." In *Gone In November* (Florastamine, 2016; 53% of 152 gamer reviews were positive), players follow the last days of a man diagnosed with depression, described as follows:

> *A short experience where your choices and actions don't matter. Texting messages to a social network account...fencing your apartment to isolate yourself from the outside world - what else someone can possible do when they are at their bottom?*

In contrast, the story of *Blue Sheep* (Noetic Games, 2016; 37% of 16 gamer reviews were positive) drew on the real experiences of the game developer:

> *Highly personal narrative driven by the developers'personal experience with depression and suicide.*

Players assume the role of the Outsider, a young black girl, who experiences "the memories of a warrior who once opposed the Beast." *Blue Sheep* is a puzzle game, in which the players overcome obstacles (a possible metaphor for life's adversities) to understand the truth behind the Beast, an adversary that symbolizes depression.

Blue Sheep, however, was exceptional in its nuanced, realistic portrayal. The majority of identified games (eg, *Reflection of Mine*, *Layers of Fear*, and *The Wendigo*) did not capture the multiplicity of emotions and lived experiences of mental illnesses like psychosis. Instead, they focused on a limited number of negative emotions such as fear and isolation and resorted to dramatizations that left little space for the dignity of those with mental illness.

Discussion

Principal Findings

In reviewing the characterization of mental illness in commercial video games, we found that the experience of mental illness, including its treatment and settings of care, was depicted in 100 games. Most of the games we reviewed (97%, 97/100) portrayed and perpetuated well-known stereotypes and prejudices associated with mental illness, namely, that those with mental illness (especially psychosis) are violent, scary, insane, abnormal, incapable, unlikely to get well, isolated, and fearful. Furthermore, some games portrayed mental illness as manifestations or consequences of supernatural phenomena or paranormal experiences. As mental illness was often associated with mystery, being unpredictable, and as an obscure illness, its treatment was also associated with uncertainties, as game characters with mental illness had to undergo "experiment treatment" to get better. Unfortunately, little or no hope for recovery was present in the identified video games, where mental illness was often presented as an ongoing struggle and an endless battle with the characters' mind and themselves.

Although mental illness can indeed be a scary and isolating experience for some, the stereotypical portrayals of mental illness that we found are problematic because they are partial, negative, and limited [47]. From unvarying characterizations of the sort we found, Adichie [48] said:

> *The single story creates stereotypes, and the problem with stereotypes is not that they are untrue, but that they are incomplete. They make one story become the only story.*

We discuss the key findings further in relation to current evidence on the impact of media portrayals of mental illness and stigma and the ability of video game technology (particularly serious video games) to promote alternative messages around mental illness and clinical practices, and we conclude by presenting future research directions for this field.

The Power of Media Portrayals and Stigma

As described, the conceptions of mental illness in the games we reviewed tended to reflect common preconceived notions of mental illnesses as dark, scary, isolated, and violent. Furthermore, the person experiencing mental illness was portrayed as having to overcome the negative experience by solitarily winning a war within oneself. The treatment of mental illness was portrayed as an experimental or an uncertain treatment, where psychiatrists were violent, insane, and dangerous. Treatment facilities, mainly represented in the form of asylums, were depicted as abandoned, disturbing, creepy, and haunted. In such settings, patients have to survive or escape to remain alive.

These findings are concerning because the promotion of predominantly negative attitudes and stereotypes is known to foster discrimination against and the marginalization of people with mental illness [49] and treatment settings (eg, mental health hospitals). This is especially salient in the context of video games because of their popularity among youth. Studies have shown that early exposure to characters stereotypically depicted as mentally ill can cause fear and anxiety in young viewers and result in them avoiding individuals with mental illness [22,37-39]. Research has also consistently demonstrated negative consequences associated with adults' stereotyped beliefs about mental illness [50,51]. Stigma and discrimination also give rise to significant barriers when people with mental illness seek help and access treatment [50,51].

Our findings align with and complement previous research that has examined the impact of TV news on attitudes toward mental illness [25,52,53]. A recent study of a random sample of 400 American news stories about mental illness published from

1995 to 2014 found that the association between mental illness and violence remained strong over time. Although 55% of the analyzed stories linked mental illness with violence (whether interpersonal or self-directed), only 14% reported on recovery from or the successful treatment of mental illness [54]. In a similar vein, research, including 1 experimental study, found that exposure to movies or news stories about mass shootings was associated with harboring negative attitudes toward persons with serious mental illness [25,52,53].

This study's findings also align with Shapiro and Rotter's findings by confirming how mental illness is often portrayed through video game characters that are violent, dysfunctional, paranoid, eccentric, afflicted, and broken [16]. Moreover, this study's findings expand on such analysis by showing that the misrepresentation of mental illness in video games is not limited to video game characters but is fully embedded into the video game experience through its environment, ambiance, and game goals.

Serious Video Games: The Potential for Alternative Messages and Clinical Practices

Co-Design Commercial Video Games for Social Good

Video games have been shown to be effective in promoting learning [55,56] because of their interactivity, repetitiveness, feedback loops, and propensity for emotional learning [57]. As such, they can provide a unique avenue for countering stigmatizing messages and "single story" narratives that create stereotypes. Video games have the potential for telling engaging stories that can shift gamers' perceptions about mental illness [56,57]. According to Bogost [58], video games have a persuasive power that goes beyond other forms of persuasion, because in video games, learning is mediated by players' in-game decisions. Bogost argues that the more interactive a game is, the more moved the player will be and, therefore, the greater the potential for intellectual persuasion.

Recently, a few game developers have taken advantage of the interactivity of games to tell engaging, real-life stories about mental illness. For instance, the developers of *Hellblade: Senua's Sacrifice* (Ninja Theory, 2017) consulted with psychiatrists and people living with psychosis to create an in-game experience that depicts psychosis realistically. Similarly, working in partnership with young gamers with lived experience of mental illness and addictions, the developers of *Debris* (Moonray Studios, 2017) also created a gaming experience that focuses on coping with psychosis during a crisis. The game demonstrates how psychosis affects not just the individual but also family members, friends, and fellow citizens [59,60].

Noteworthy is the fact that *Debris* was designed using participatory methodology. Its design process involved the input and perspectives of 5 youth with lived experience of mental illness, Moonray Studio's founder and project manager, and academic researchers. These stakeholders met for 8 months to discuss video game elements (game characters, story, interactivity, graphics, etc), video game experiences, and messages around psychosis and mental illness [61]. The 5 youth felt that serious video games that attempt to describe the experience of living with mental illness should express empathy, dignity, and compassion. Guided by these values, *Debris* sought to promote messages of compassion and support toward mental illness, rather than fear.

In addition to being developed in consultation with persons with lived experience of and professionals with expertise in mental illness, *Hellblade: Senua's Sacrifice* and *Debris* share certain attributes. Both games acknowledge the link between traumatic experiences and psychosis, demonstrate how illness symptoms (eg, hearing voices) are not always threatening or perceived negatively by persons with psychosis, and depict the different ways in which psychosis can be experienced. They exemplify how video games can employ features, mechanisms, characters, and story lines that can disrupt mental illness stereotypes and potentially contribute significant new understandings of the illness experience.

From the Console to the Clinic: Video Games for Clinical Treatment

With growing acknowledgment of their potential for positively impacting learning, video games are now used in treatment settings to promote symptom remission and recovery from mental illness. In a 2012 study, the video game *SPARX* was found to effectively reduce depressive symptoms among adolescents aged 12 to 19 years [62]. The authors concluded that *SPARX* was a potential alternative to usual care for adolescents with depressive symptoms in primary care settings.

Similarly, the game *MindLight* (Playnice Institute) was effective in reducing anxiety symptoms among schoolchildren and children with autism spectrum disorder [63]. The mechanisms deployed in *MindLight* include *exposure techniques* (empirically validated treatment components of cognitive behavioral therapy for anxiety) whereby players are gradually exposed and habituated to threatening cues until they become comfortable with and less anxious about them; *neurofeedback mechanisms* used to promote self-monitoring, relaxation, and concentration; and *attention bias modification* that teaches players to shift attention away from threatening cues and focus on positive aspects of the environment to the end of attaining relevant goals.

Developed for youth and young adults, *Reach Out Central* is a game that enables players to learn skills and information about health-seeking behaviors and progress to higher levels based on the skill and knowledge achieved [64].

Such new-generation, evidence- and lived experience–informed video games can effectively motivate and facilitate mental health learning. They can even open possibilities for envisioning new modalities for providing care to youth and supporting caregivers and health care providers [62-65].

Limitations

Being based solely on one game platform (Steam) for PC games, our search engine did not capture all available console games (PlayStation, Xbox, etc). Having limited ourselves for feasibility concerns to games published over the course of 18 months, we could not capture the evolution of messages about mental illness in video games over any significant length of time.

Furthermore, in this review, we focused our keywords on psychosis because it is known to be the mental illness that is most frequently and strongly stigmatized [66,67]. Moreover, our list of words was not exhaustive, as words such as “nuts” were not included. Despite not including “depression” as a keyword, we did identify some games with depression. It is possible that some games may not have been captured by the keywords we selected. Finally, not having played the games identified, we could not comprehensively assess how mental illness was depicted in them.

In addition, this review does not fully assess if any of the identified games were troll games or scam games that are posted to generate revenue but do not satisfy game design value or players’ needs. Finally, we recognized that the impact that the 100 reviewed games have on the market is quite heterogeneous. If we look at the game analytics, some games were reviewed by a small number of players, whereas others by more than 15,000 gamers. Game players’ comments can be used to assess the impact of the game; unfortunately, it is a poor measure as not all gamers post a review and gamer satisfaction is based on many different elements (eg, graphics, game experience). Despite this acknowledgment of the variable reach the games reviewed have, this review begins to unpack the negative portrayals of mental illness by game designers and video game industries.

These limitations notwithstanding, we employed novel methodological and analytical approaches to explore mental illness messages in video games. Video games are a new form of media that has yet to be fully examined in the literature base that has hitherto focused primarily on the depiction of mental illness in TV, news reports, and movies. This study’s methods were similar but also different from methods described by Shapiro and Rotter [16]. Our search was comprehensive but also focused, as we looked at all video games on Steam between January 2016 and June 2017, without the use of a predefined analytical framework, as Shapiro and Rotter did. We instead opted for an inductive approach. Finally, we expanded our analysis to assess the representation of mental illness through different game elements. As such, this review contributes significantly to filling an important knowledge gap. Its salience lies in the popularity of video games among young people, and the strong potential for video games in enhancing learning through repetition, interactivity, and immersion.

Future Directions

Future research initiatives can work to validate the themes identified in this review. An ethnographic study is already underway on the gaming experience of the games identified in this review. Further research exploring players’ perspectives on these games and how they portray mental illness is warranted. Finally, an experimental study is required to assess how and how much the messages about mental illness that this review has identified shape players’ knowledge of, attitudes and behavior toward, and beliefs about mental illness and persons with mental illness.

Conclusions

The video game industry and its consumers need to be educated about the potential negative impact of ill-conceived messages about mental illness and how these stereotypes can drive discriminatory behavior. As a new generation of collaboratively developed games have shown, much stands to be gained from researchers and clinicians partnering with the gaming industry to create games that can contain and promote positive, nuanced, realistic, and compelling messages about mental illness. Such games can meet gamers’ desires for adventure, pleasure, challenge, and esthetics (eg, *Debris*) as well as meet the video game industry’s desire to produce a successful and popular video game, such as *Hellblade: Senua’s Sacrifice*, which sold more than 1 million copies within a few months and had 91% of 17,032 gamers rating the game as “Very Positive” on Steam.

Acknowledgments

SM received funds to support her graduate studies from the Department of Psychiatry at McGill University, Fonds de Recherche du Québec-Santé (FRQS), ACCESS Open Minds (a pan-Canadian youth mental health services research network), and the Canadian Institutes of Health Research (CIHR). GJ received funds to support his graduate studies from the Department of Psychiatry at McGill University, FRQS, and CIHR. JS has received funding from FRQS. SL is supported by the Canada Research Chairs Program. SNI is supported by a salary award from CIHR and has a grant from the CIHR, which supported this research.

Authors' Contributions

MF contributed to the project conception, methodological design, data collection, analysis and writing of this manuscript. SM has contributed to the data collection, analysis, and writing of this manuscript. GJ, JS, SL, and SNI have contributed to the writing of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary table of game analytics.

[PDF File (Adobe PDF File), 253KB - mental_v6i4e12418_app1.pdf]

References

1. Entertainment Software Association. 2017. 2017 Essential facts about the computer and video game industry URL: http://www.theesa.com/article/2017-essential-facts-computer-video-game-industry/ [accessed 2019-04-08] [WebCite Cache ID 77Tyq1itb]
2. Lenhart A, Kahne J, Middaugh E, Macgill A, Evans C, Vitak J. Pew Internet. 2008. Teens, Video Games, and Civics: Teens' Gaming Experiences Are Diverse and Include Significant Social Interaction and Civic Engagement URL: https://files.eric.ed.gov/fulltext/ED525058.pdf [accessed 2019-04-09] [WebCite Cache ID 77VX0AdJN]
3. Anderson CA, Bushman BJ. Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: a meta-analytic review of the scientific literature. Psychol Sci 2001 Sep;12(5):353-359. [doi: 10.1111/1467-9280.00366] [Medline: 11554666]
4. Gentile D. Pathological video-game use among youth ages 8 to 18: a national study. Psychol Sci 2009 May;20(5):594-602. [doi: 10.1111/j.1467-9280.2009.02340.x] [Medline: 19476590]
5. Zendle D, Kudenko D, Cairns P. Behavioural realism and the activation of aggressive concepts in violent video games. Entertain Comput 2018 Jan;24:21-29. [doi: 10.1016/j.entcom.2017.10.003]
6. Granic I, Lobel A, Engels RC. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
7. Colder Carras M, Van Rooij AJ, Spruijt-Metz D, Kvedar J, Griffiths MD, Carabas Y, et al. Commercial video games as therapy: a new research agenda to unlock the potential of a global pastime. Front Psychiatry 2018 Jan 22;8:300 [FREE Full text] [doi: 10.3389/fpsyt.2017.00300] [Medline: 29403398]
8. Barnes S, Prescott J. Empirical evidence for the outcomes of therapeutic video games for adolescents with anxiety disorders: systematic review. JMIR Serious Games 2018 Feb 28;6(1):e3 [FREE Full text] [doi: 10.2196/games.9530] [Medline: 29490893]
9. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. Front Psychiatry 2017 Jan 18;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
10. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2017 Jan 10;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
11. Ashmore R, Del Boca F. Conceptual approaches to stereotypes and stereotyping. In: Hamilton DL, editor. Cognitive Processes in Stereotyping and Intergroup Behavior. Hillsdale, NJ: Lawrence Erlbaum; 1981.
12. Dates JL, Barlow W. Split Image: African Americans in the Mass Media. Washington, DC: Howard University Press; 1993.
13. Beasley B, Collins Standley T. Shirts vs skins: clothing as an indicator of gender role stereotyping in video games. Mass Commun Soc 2002 Aug;5(3):279-293. [doi: 10.1207/S15327825MCS0503_3]
14. Jenson J, de Castell S. Gender, simulation, and gaming: research review and redirections. Simul Gaming 2010 Jan 26;41(1):51-71. [doi: 10.1177/1046878109353473]
15. Sendler DJ. Commentary on: Shapiro S, Rotter M. Graphic depictions: portrayals of mental illness in video games. J Forensic Sci 2016;61(6):1592-5. J Forensic Sci 2017 May;62(3):831. [doi: 10.1111/1556-4029.13482] [Medline: 28449260]
16. Shapiro S, Rotter M. Graphic depictions: portrayals of mental illness in video games. J Forensic Sci 2016 Nov;61(6):1592-1595. [doi: 10.1111/1556-4029.13214] [Medline: 27783392]
17. Alves T. University of Scovde. 2018. Exploring Underrepresented Narrativesocial Anxiety in Games URL: http://www.diva-portal.org/smash/get/diva2:1217926/FULLTEXT01.pdf [accessed 2019-04-08] [WebCite Cache ID 77U0m5d0e]
18. Hyler SE, Gabbard GO, Schneider I. Homicidal maniacs and narcissisfic parasites: stigmatization of mentally ill persons in the movies. Psychiatr Serv 1991 Oct;42(10):1044-1048. [doi: 10.1176/ps.42.10.1044]
19. Angermeyer MC, Matschinger H. The effect of violent attacks by schizophrenic persons on the attitude of the public towards the mentally ill. Soc Sci Med 1996 Dec;43(12):1721-1728. [doi: 10.1016/S0277-9536(96)00065-2]
20. Goulden R, Corker E, Evans-Lacko S, Rose D, Thornicroft G, Henderson C. Newspaper coverage of mental illness in the UK, 1992-2008. BMC Public Health 2011 Oct 12;11:796-748 [FREE Full text] [doi: 10.1186/1471-2458-11-796] [Medline: 21992410]
21. Wahl OF. Mass media images of mental illness: a review of the literature. J Community Psychol 1992 Oct;20(4):343-352. [doi: 10.1002/1520-6629(199210)20:4<343::AID-JCOP2290200408>3.0.CO;2-2]
22. Wilson C, Nairn R, Coverdale J, Panapa A. How mental illness is portrayed in children's television. Br J Psychiatry 2000 May;176(5):440-443. [doi: 10.1192/bjp.176.5.440]
23. Lawson A, Fouts G. Mental illness in Disney animated films. Can J Psychiatry 2004 May;49(5):310-314. [doi: 10.1177/070674370404900506] [Medline: 15198467]
24. Signorielli N. The stigma of mental illness on television. J Broadcast Electron Media 1989 Jun;33(3):325-331. [doi: 10.1080/08838158909364085]
25. Wahl OF, Lefkowits JY. Impact of a television film on attitudes toward mental illness. Am J Community Psychol 1989 Aug;17(4):521-528. [Medline: 2610206]

26. Domino G. Impact of the film, "One Flew Over the Cuckoo's Nest," on attitudes towards mental illness. Psychol Rep 1983 Aug;53(1):179-182. [doi: 10.2466/pr0.1983.53.1.179] [Medline: 6635061]
27. Shapiro J. Freeplay: A video game guide to maximum euphoric bliss. Scotts Valley, California: CreateSpace Independent Publishing Platform; 2012.
28. Rosenstock J. Beyond a beautiful mind: film choices for teaching schizophrenia. Acad Psychiatry 2003;27(2):117-122. [doi: 10.1176/appi.ap.27.2.117] [Medline: 12824113]
29. Owen PR. Portrayals of schizophrenia by entertainment media: a content analysis of contemporary movies. Psychiatr Serv 2012 Jul;63(7):655-659. [doi: 10.1176/appi.ps.201100371] [Medline: 22555313]
30. Fairclough N. Analysing Discourse: Textual Analysis For Social Research. London: Routledge; 2019.
31. Link B, Phelan J. Labeling and stigma. In: Aneshensel CS, Phelan JC, Bierman A, editors. Handbooks of Sociology and Social Research. New York: Springer Science & Business Media; 2013:525-541.
32. Swanson JW, McGinty EE, Fazel S, Mays VM. Mental illness and reduction of gun violence and suicide: bringing epidemiologic research to policy. Ann Epidemiol 2015 May;25(5):366-376 [FREE Full text] [doi: 10.1016/j.annepidem.2014.03.004] [Medline: 24861430]
33. Baumann AE. Stigmatization, social distance and exclusion because of mental illness: the individual with mental illness as a 'stranger'. Int Rev Psychiatry 2007 Apr;19(2):131-135. [doi: 10.1080/09540260701278739] [Medline: 17464791]
34. Link BG, Struening EL, Rahav M, Phelan JC, Nuttbrock L. On stigma and its consequences: evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. J Health Soc Behav 1997 Jun;38(2):177. [doi: 10.2307/2955424]
35. Stuber J, Meyer I, Link B. Stigma, prejudice, discrimination and health. Soc Sci Med 2008 Aug;67(3):351-357 [FREE Full text] [doi: 10.1016/j.socscimed.2008.03.023] [Medline: 18440687]
36. Wahl OF. Mental health consumers' experience of stigma. Schizophr Bull 1999 Jan 01;25(3):467-478. [doi: 10.1093/oxfordjournals.schbul.a033394]
37. Adler AK, Wahl OF. Children's beliefs about people labeled mentally ill. Am J Orthopsychiatry 1998;68(2):321-326. [doi: 10.1037/h0080341]
38. Weiss MF. Children's attitudes toward the mentally ill: a developmental analysis. Psychol Rep 1986 Feb;58(1):11-20. [doi: 10.2466/pr0.1986.58.1.11] [Medline: 3961053]
39. Weiss MF. Children's attitudes toward the mentally ill: an eight-year longitudinal follow-up. Psychol Rep 1994 Feb;74(1):51-56. [doi: 10.2466/pr0.1994.74.1.51] [Medline: 8153235]
40. Papadopoulos J. 2017. PC games sales in 2017 are almost as big as all console sales combined URL: https://www.dsogaming.com/news/pc-games-sales-in-2017-are-almost-as-big-all-console-sales-combined/ [accessed 2019-04-08] [WebCite Cache ID 77U28s5y8]
41. Atlas.ti: Qualitative Data Analysis Software. Berlin: Scientific Software Development; 2013. URL: https://atlasti.com [accessed 2019-04-09] [WebCite Cache ID 77VW7fHKx]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Willig C. Discourse analysis. In: Smith JA, editor. Qualitative Psychology: A Practical Guide to Research Methods, Volume 2. Thousand Oaks, California: Sage Publications; 2003.
44. Penn D, Nowlin-Drummond A. Politically correct labels and schizophrenia: a rose by any other name? Schizophr Bull 2001;27(2):197-203. [doi: 10.1093/oxfordjournals.schbul.a006866] [Medline: 11354587]
45. Imhoff R. Zeroing in on the effect of the schizophrenia label on stigmatizing attitudes: a large-scale study. Schizophr Bull 2016 Mar;42(2):456-463 [FREE Full text] [doi: 10.1093/schbul/sbv137] [Medline: 26409222]
46. Dickey M. Engaging by design: how engagement strategies in popular computer and video games can inform instructional design. Educ Technol Res Dev 2005 Jun;53(2):67-83. [doi: 10.1007/BF02504866]
47. Sieff E. Media frames of mental illnesses: the potential impact of negative frames. J Ment Health 2009 Jul 6;12(3):259-269. [doi: 10.1080/0963823031000118249]
48. Adichie C. National Geographic. 2009. The danger of a single story URL: https://ngl.cengage.com/21centuryreading/resources/sites/default/files/B3_TG_AT7_0.pdf [accessed 2019-04-08] [WebCite Cache ID 77U2qHC1S]
49. Corrigan P, Kleinlein P. The impact of mental illness stigma. In: Corrigan PW, editor. On the Stigma of Mental Illness: Practical Strategies for Research and Social Change. Washington, DC: American Psychological Association; 2005:11-44.
50. Angermeyer M, Dietrich S. Public beliefs about and attitudes towards people with mental illness: a review of population studies. Acta Psychiatr Scand 2006 Mar;113(3):163-179. [doi: 10.1111/j.1600-0447.2005.00699.x] [Medline: 16466402]
51. Link B, Struening E, Neese-Todd S, Asmussen S, Phelan J. Stigma as a barrier to recovery: the consequences of stigma for the self-esteem of people with mental illnesses. Psychiatr Serv 2001 Dec;52(12):1621-1626. [doi: 10.1176/appi.ps.52.12.1621] [Medline: 11726753]
52. McGinty EE, Webster DW, Barry CL. Effects of news media messages about mass shootings on attitudes toward persons with serious mental illness and public support for gun control policies. Am J Psychiatry 2013 May;170(5):494-501. [doi: 10.1176/appi.ajp.2013.13010014] [Medline: 23511486]

53. Thornton J, Wahl O. Impact of a newspaper article on attitudes toward mental illness. J Community Psychol 1996 Jan;24(1):17-25. [doi: 10.1002/(SICI)1520-6629(199601)24:1<17::AID-JCOP2>3.0.CO;2-0]
54. McGinty EE, Kennedy-Hendricks A, Choksy S, Barry CL. Trends in news media coverage of mental illness in the United States: 1995-2014. Health Aff (Millwood) 2016 Dec 1;35(6):1121-1129 [FREE Full text] [doi: 10.1377/hlthaff.2016.0011] [Medline: 27269031]
55. Gee J. Learning by design: good video games as learning machines. E-learn Digit Media 2016 Jul 31;2(1):5-16. [doi: 10.2304/elea.2005.2.1.5]
56. Shaffer DW, Squire KR, Halverson R, Gee JP. Video games and the future of learning. Phi Delta Kappan 2005 Oct;87(2):105-111. [doi: 10.1177/003172170508700205]
57. Wilson KA, Bedwell WL, Lazzara EH, Salas E, Burke CS, Estock JL, et al. Relationships between game attributes and learning outcomes: review and research proposals. Simul Gaming 2008 May 7;40(2):217-266. [doi: 10.1177/1046878108321866]
58. Bogost I. Persuasive Games: The Expressive Power of Videogames. Cambridge, MA: MIT Press; 2007.
59. Carter A. CBC News. 2017. New video game shows players what it's like to experience psychosis URL: https://www.cbc.ca/news/canada/hamilton/debris-game-1.4226303 [accessed 2019-04-08] [WebCite Cache ID 77U3dsWLe]
60. Ferrari M, Bush N, Clark D, Archie S. DIGRA. 2016. Debris: Exploring the video game values that can help reduce mental illness stigma URL: http://www.digra.org/digital-library/publications/debris-exploring-the-video-game-values-that-can-help-reduce-mental-illness-stigma/ [accessed 2019-04-08] [WebCite Cache ID 77U3sdOre]
61. Ferrari M, Archie S, Bush N, Clark D. DIGRA. 2018. Debris: Exploring video game messages and values through gameplay URL: http://www.digra.org/digital-library/publications/debris-exploring-video-game-messages-and-values-through-gameplay/ [accessed 2019-04-09] [WebCite Cache ID 77VRgadtp]
62. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. Br Med J 2012 Apr 19;344(apr18 3):e2598. [doi: 10.1136/bmj.e2598]
63. Schoneveld E, Malmberg M, Lichtwarck-Aschoff A, Verheijen G, Engels R, Granic I. A neurofeedback video game (MindLight) to prevent anxiety in children: a randomized controlled trial. Comput Human Behav 2016 Oct;63:321-333. [doi: 10.1016/j.chb.2016.05.005]
64. Shandley K, Austin D, Klein B, Kyrios M. An evaluation of 'Reach Out Central': an online gaming program for supporting the mental health of young people. Health Educ Res 2010 Aug;25(4):563-574 [FREE Full text] [doi: 10.1093/her/cyq002] [Medline: 20150530]
65. Wijnhoven LA, Creemers DH, Engels RC, Granic I. The effect of the video game Mindlight on anxiety symptoms in children with an Autism Spectrum Disorder. BMC Psychiatry 2015 Jul 1;15:138 [FREE Full text] [doi: 10.1186/s12888-015-0522-x] [Medline: 26129831]
66. Rose D, Thornicroft G, Pinfold V, Kassam A. 250 labels used to stigmatise people with mental illness. BMC Health Serv Res 2007 Jun 28;7:97 [FREE Full text] [doi: 10.1186/1472-6963-7-97] [Medline: 17598894]
67. Thornicroft G, Brohan E, Rose D, Sartorius N, Leese M, INDIGO Study Group. Global pattern of experienced and anticipated discrimination against people with schizophrenia: a cross-sectional survey. Lancet 2009 Jan 31;373(9661):408-415. [doi: 10.1016/S0140-6736(08)61817-6] [Medline: 19162314]

Abbreviations

CIHR: Canadian Institutes of Health Research
FRQS: Fonds de Recherche du Québec-Santé
PC: personal computer
TV: television

Edited by J Torous, M Birk; submitted 05.10.18; peer-reviewed by C Ball, L Bonet; comments to author 26.10.18; revised version received 25.01.19; accepted 29.01.19; published 08.05.19.

Please cite as:
Ferrari M, McIlwaine SV, Jordan G, Shah JL, Lal S, Iyer SN
Gaming With Stigma: Analysis of Messages About Mental Illnesses in Video Games
JMIR Ment Health 2019;6(5):e12418
URL: https://mental.jmir.org/2019/5/e12418/
doi:10.2196/12418
PMID:31066703

Original Paper

Using Computer Games to Support Mental Health Interventions: Naturalistic Deployment Study

Hidde van der Meulen[1], BEng, MSc; Darragh McCashin[2], BA, MSc; Gary O'Reilly[2], PhD; David Coyle[1], PhD

[1]University College Dublin, School of Computer Science, Dublin, Ireland

[2]University College Dublin, School of Psychology, Dublin, Ireland

Corresponding Author:
David Coyle, PhD
University College Dublin
School of Computer Science
Belfield
Dublin, 4
Ireland
Phone: 353 17162818
Email: d.coyle@ucd.ie

Abstract

Background: Recent research has highlighted *naturalistic uptake* as a key barrier to maximizing the impact of mental health technologies. Although there is increasing evidence regarding the efficacy of digital interventions for mental health, as demonstrated through randomized controlled trials, there is also evidence that technologies do not succeed as expected when deployed in real-world settings.

Objective: This paper describes the naturalistic deployment of Pesky gNATs, a computer game designed to support cognitive behavioral therapy (CBT) for children experiencing anxiety or low mood. The objective of this deployment study was to identify how therapists use Pesky gNATs in real-world settings and to discover positive and negative factors. On the basis of this, we aimed to derive generalizable recommendations for the development of mental health technologies that can have greater impact in real-world settings.

Methods: Pesky gNATs has been made available through a not-for-profit organization. After 18 months of use, we collected usage and user experience data from therapists who used the game. Data were collected through an online survey and semistructured interviews addressing the expectations and experiences of both therapists and young people. Thematic analysis was used to identify key themes in the interview and survey data.

Results: A total of 21 therapists, who used Pesky gNATs with 95 young people, completed the online survey. Furthermore, 5 therapists participated in the follow-up interview. Confirming previous assessments, data suggest that the game can be helpful in delivering therapy and that young people generally liked the approach. Therapists shared diverse opinions regarding the young people for whom they deemed the game appropriate. The following 3 themes were identified: (1) stages of use, (2) impact on the delivery of therapy, and (3) customization. We discuss therapists' reflections on the game with regard to their work practices and consider the question of customization, including the delicate balance of adaptable interaction versus the need for fidelity to a therapeutic model.

Conclusions: This study provides further evidence that therapeutic games can support the delivery of CBT for young people in real-world settings. It also shows that deployment studies can provide a valuable means of understanding how technologies integrate with the overall mental health ecosystem and become a part of therapists' toolbox. Variability in use should be expected in real-world settings. Effective training, support for therapist autonomy, careful consideration of different approaches to customization, the reporting of deployment data, and support for communities of practice can play an important role in supporting variable, but effective, use.

(JMIR Ment Health 2019;6(5):e12430) doi:10.2196/12430

KEYWORDS

mental health; eHealth; mHealth; adolescents; children; computer game; treatment; cognitive behavioral therapy

Introduction

Youth Mental Health

Recent years have witnessed significant progress in technologies designed to support improved mental health in adults [1-5]. Although the evidence on the potential of technology to support interventions with young people is growing [6,7], the body of research in this area is more limited. This stands in sharp contrast to the stated objectives of the international health policies, where the prevention and treatment of mental health problems is a central objective [8,9], and where young people are identified as particularly vulnerable [10] and requiring specific attention [11,12]. Evidence suggests that 50% of mental disorders emerge by 14 years of age and 75% present before 24 years of age [13]. The Global Burden of Disease figures reveal that young people aged between 10 and 24 years represent 27% of the world's population within which poor mental health is the leading cause of disability [14]. Unaddressed mental health difficulties in early life are associated with at least a threefold increase in the odds of having a mental health disorder in later life [15]. Early intervention is widely encouraged to save lives, prevent young people from falling into crisis, and help in reducing the need for expensive long-term interventions in adulthood [13]. Patel et al argue that "more research is urgently needed to improve the range of affordable and feasible interventions, as most mental-health needs in young people are unmet, even in high-income countries" [10].

Therapeutic Games

This paper focuses on Pesky gNATs, a computer game designed to support the delivery of cognitive behavioral therapy (CBT) for young people experiencing low mood or anxiety [16]. Pesky gNATs is located among a growing body of research on the use of serious games in mental health [17-19]. Other examples include SPARX [20], Rainbow SPARX [21], The Feel Good Island [22], Op Volle Kracht [23], Personal Investigator [24], Ricky the Spider [25], Treasure Hunt [26], and The Journey [27]. Many of these games have been the subject of small-scale evaluations. Pesky gNATs itself builds on a previous game, gNATs Island (see evaluation in [28]). However, larger-scale studies have also been conducted. A randomized controlled noninferiority trial of SPARX demonstrated that it was a viable alternative to usual care for adolescents presenting with depressive symptoms in primary care. By reducing the costs and clinical resources associated with treatment, it has the potential to help in addressing unmet demand [29]. A randomized controlled trial (RCT) of The Feel Good Island demonstrated the potential of a serious game to support CBT for adults with intellectual disabilities [22].

Naturalistic Evaluations of Digital Interventions

Pesky gNATs is currently being evaluated through a multisite RCT in primary care psychology services in Ireland [30]. That study will allow us to understand the controlled use of Pesky gNATs and provide robust evidence on the efficacy of the game. This paper explores a different approach, focusing on the use of the game in real-world settings, that is, outside of controlled settings. We believe this approach is complementary to traditional RCTs and can help maximize the impact of digital interventions.

The importance of alternative evaluation strategies for digital interventions, including *naturalistic* studies, is increasingly recognized by researchers [18,31,32]. In part, this reflects a misalignment between the timescales of RCTs and the rapid advancement and subsequent obsolescence of technologies. It also reflects the limitations of traditional RCTs in understanding the impact of digital interventions when deployed in real-world contexts. Discussing the evolving, interdisciplinary nature of research on digital interventions, Blandford et al argue that "If interactive digital health interventions are to be sustainable, and a good investment...then it is essential to invest time and effort in ensuring that they are incorporated within broader care delivery systems in ways that work well for professionals, patients, and others involved in care" [33]. Increasingly, researchers argue that implementation data should be routinely analyzed to assess technologies and their fit in the real-world [32,34].

These arguments are part of the ongoing discussions within the literature on how to optimize the impact of digital technologies within day-to-day services, with Fleming et al suggesting that mental health technologies are not reaching or engaging the intended users post release [34]. A comparison of real-world usage and that found in controlled studies of MoodGYM, found that adherence of users in the RCT [35] was significantly higher than adherence in public users [36]. This is particularly important given the evident correlation between adherence and efficacy [37]. Another large-scale RCT assessing MoodGYM and Beating the Blues also found unexpected results. In contrast with previous RCTs, no significant differences were found between the digital interventions and usual care [38].

Similar discrepancies can be found within serious games literature. Poppelaars et al [39] described a school-based trial with adolescents reporting subclinical depressive symptoms in which they compared the games Op Volle Kracht, SPARX, and a combination of the 2 with a wait-list group. The study showed no significant differences of depressive symptoms among the 4 groups [39]. The authors found comparable results in other studies about Op Volle Kracht but found conflicting results with a similar study by Wijnhoven et al, which reported positive outcomes for the program [40]. Poppelaars et al suggested that the differences may be attributable to stricter study inclusion criteria used by Wijnhoven et al. In a similar comparison regarding SPARX, contrasting results were found between the studies by Poppelaars et al [39] and Fleming et al [20]. It is argued that these differences can be caused by the different participant environments (school-going vs school-excluded). These findings demonstrated that small differences in study criteria can have a significant impact on outcomes.

As noted above, this paper is grounded in the naturalistic deployment of Pesky gNATs. The aim of this deployment was to identify how therapists used Pesky gNATs in real-world settings. Building on this, the broader aim of this paper was to contribute toward generalizable recommendations for the design of mental health technologies that can have greater impact in real-world settings.

Pesky gNATs

Pesky gNATs is a 3-dimensional (3D) computer game that helps to teach young people CBT concepts and skills [16,41]. Here we briefly outlined the key aspects of the game and its theoretical basis. The CBT content of Pesky gNATs is further described in Multimedia Appendix 1, with a more detailed description available in the study by O'Reilly [41].

Pesky gNATs is played by young people with clinically significant anxiety or low mood, during face-to-face therapy sessions with a suitably qualified mental health professional. In sessions, the young person and therapist sit together at a computer, but the young person has full control of the mouse and keyboard. They navigate through a 3D world where they meet a series of characters who introduce CBT concepts using spoken conversation, embedded animations, videos, and questions regarding the young person's own situation (Figure 1). The in-session computer game is supplemented by a mobile phone app that aims to support young people in transferring what they learn in therapy to their home, school, and day-to-day life [16].

In addition to the game and app, the overall program includes 7 hours of online training videos for therapists. This provides a description of the concepts in the program and a full walk-through of the game and app, illustrating how they are typically used during therapy. Our intention with Pesky gNATs was to assist mental health professionals in delivering a genuinely *cognitive* CBT intervention, filtered through the ideas of developmental psychology and learning theory [42-46], packaged in a computer game that is played within the supportive context of a therapeutic relationship.

The gNATs of the game's title is a play on the CBT concept of negative automatic thoughts. Each level of the game introduces young people to a single CBT concept, which is explained and illustrated by a game character. The child, with assistance from the therapist, applies the CBT concept to the anxiety or low mood difficulties they are experiencing. The game characters set the young person a challenge of putting that concept into practice in his or her daily life. The young person then has the opportunity to learn 1 of 9 behavioral activation, relaxation, or mindfulness skills chosen by the therapist to best meet his or her current needs. Each game level takes about 50 min to play, which is the typical length of standard child therapy appointments. Completion of between session tasks is supported through the Pesky gNATs app or a hard copy workbook. Overall, Pesky gNATs is intended to provide a coherent and consistently delivered child-friendly introduction to CBT through its programmed content, combined with an individually nuanced, personal delivery supported by the judgement and accompanying conversations young people have with their therapists.

Figure 1. The player controls a character, exploring an island (left), and meets characters who explain cognitive behavioral therapy using comprehensible metaphors. The player also meets Ben, the Beach Dude (right), who teaches mindfulness and relaxation skills.

Sit comfortably in your chair and you can keep your eyes open. This exercise is called "Sounds Around".

Methods

Overview

This study focused on therapists' experiences of using Pesky gNATs during the first 18 months of deployment. The software and online training materials had been made available on a not-for-profit basis on a website [16]. Before accessing the software, therapists were required to complete an online registration process that included a screening questionnaire. The questionnaire was manually screened by GOR to determine if applicants had appropriate qualifications in the delivery of CBT for young people. Therapists meeting the requirements received a username and password that enabled them to access a secured website, where they could purchase a license for Pesky gNATs, download the software, and access the training materials.

Data Collection and Analysis

Data were collected through an online survey and optional semistructured interviews with therapists. The survey was implemented using LimeSurvey. All submissions were anonymous unless participants opted in for the follow-up interview, in which case contact details were required. Qualitative and quantitative data were collected with the aim of capturing how Pesky gNATs was used, identifying positive and negative factors, and exploring the expectations and experiences of therapists and young people. The survey was divided across the following topics: (1) use of Pesky gNATs, (2) helpfulness of Pesky gNATs, (3) therapy style, (4) pacing of Pesky gNATs, (5) game design, (6) customization, and (7) future developments. The inclusion of specific questions on customization reflected arguments that modularization or customization might be beneficial in supporting naturalistic uptake of digital interventions [47].

Follow-up semistructured interviews were conducted via phone or video call and followed the remote interview procedures described by Braun and Clarke [48]. Interviews addressed similar topics to those in the online survey but allowed us to explore participants' attitudes and experiences in greater detail. HM conducted and audio recorded the interviews, which lasted 40 to 60 min. Verbatim transcriptions of the recordings were made. Each transcript was manually checked for accuracy.

As the topics of the open questions in the survey had a strong overlap with those discussed in the semistructured interviews, we analyzed and presented the qualitative results from both sources together. Thematic analysis was used to identify codes and themes across the entire dataset, following the guidelines described by Braun and Clarke [48]. Furthermore, 10% of the coded data were screened by a different researcher (DMC), and disagreements were discussed until consensus was reached.

Ethical approval for this study was granted by the Research Ethics Committee at University College Dublin (reference LS-E-17-150-VanderMeulen-Coyle). Survey participants gave consent before the first question, and written consent was obtained before the start of the interviews.

Participants

An invitation to participate in the online survey was emailed to all Pesky gNATs license holders. In total, 91 survey invitations were sent. A total of 21 participants completed the survey with 5 opting in for follow-up interviews. On average, survey participants had a Pesky gNATs license for 429 days (SD 257 days) at the point of data collection (June 2018). In total, they had started the program with 95 young people (mean young participants per therapist=4.5 [SD 4.2]), ranging from 1 to 16 children per therapist. They estimated that they had completed the full 7 session program with 39 children (mean=1.9 [SD 2.6]), ranging from 0 to 12 programs completed per therapist.

Results

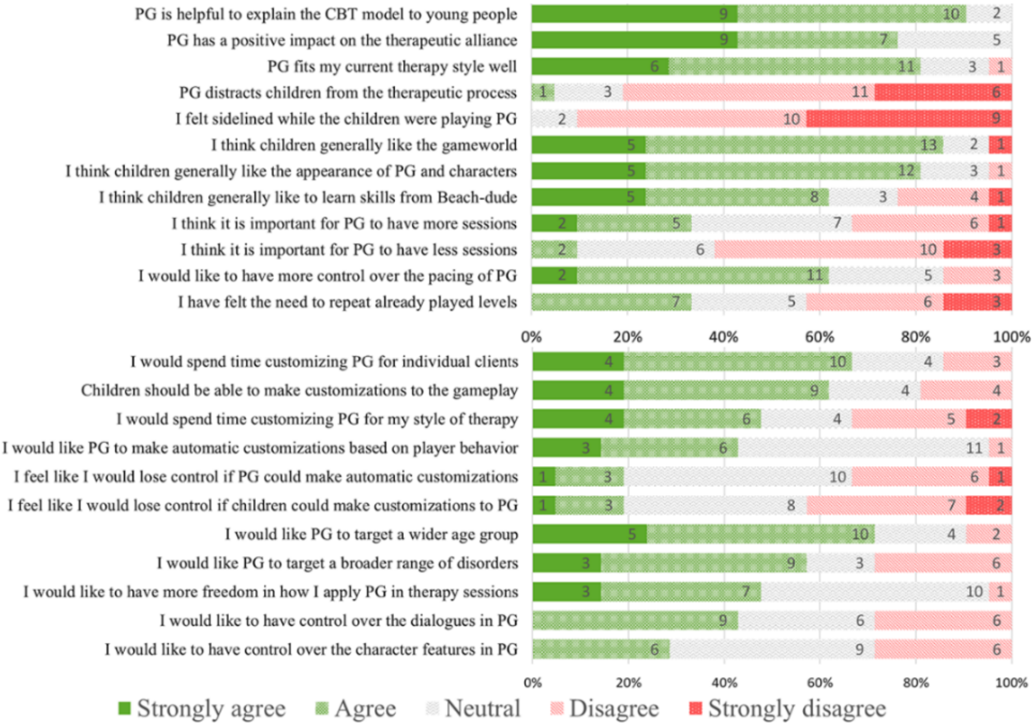
Overall Opinions

On the basis of their experiences in delivering therapy with Pesky gNATs, therapists were asked to estimate the age range for which they deemed the program most appropriate. Responses varied from 6 to 13 years as the minimum age (mean=8.8 [SD 1.5] years) and 10 to 21 years as the maximum age (mean=14.0 [SD 2.5] years). In addition, estimations of the play time for individual Pesky gNATs sessions ranged from 15 to 60 min (mean=36 [SD 11]).

Likert questions captured the therapists' overall opinions on the helpfulness of the game, children's experience, pacing of the game, and attitudes regarding different approaches to customization. The aggregated responses are shown in Figure 2.

Overall, the therapists rated the helpfulness of Pesky gNATs positively, suggesting that it helped to explain the CBT model to young people and indicating that it fitted well with their therapy style. They indicated that the program had a positive impact on the therapeutic alliance. It was generally considered not to distract from the therapeutic process, and the therapists did not feel sidelined when children controlled the game. Therapists felt that the children generally liked the game world and characters and responded well to exercises introduced by game characters.

Figure 2. Therapists' (n=21) responses to Likert questions about Pesky gNATs, its fit to their therapeutic practice, estimation of the children's experiences, and attitudes related to the pacing, customization, and target group. Numbers represent absolute number of occurrences. CBT: cognitive behavioral therapy; PG: Pesky gNATs.

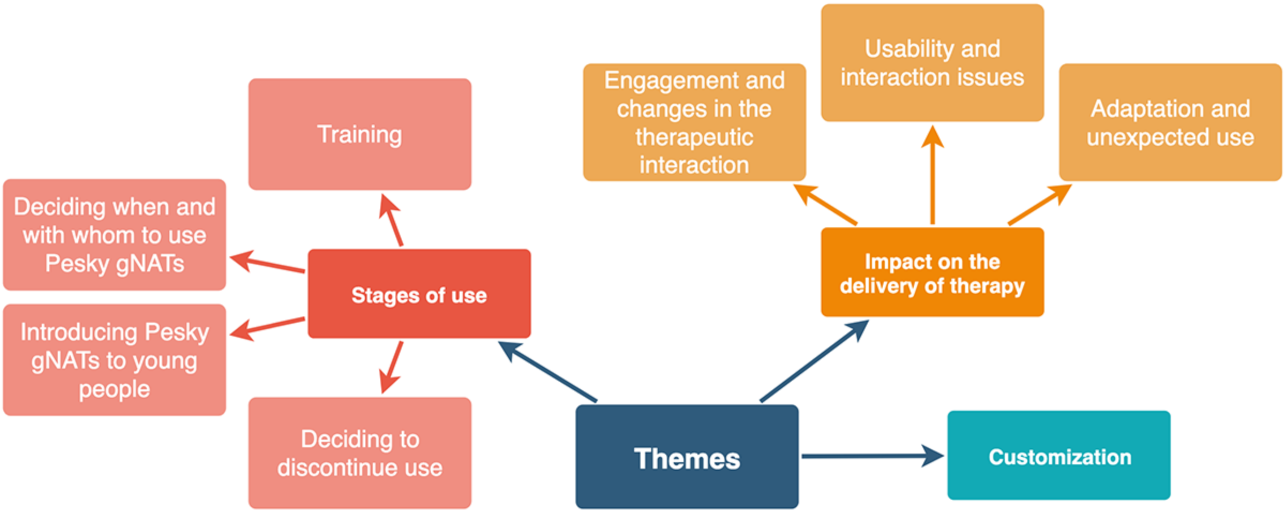


Overall, the therapists expressed positive attitudes regarding the potential for customizations, indicating that they would spend time making customizations to support individual clients. They also indicated a preference to see the game extended to target a wider age group and were generally positive about the idea of the game targeting a broader range of disorders. Although more positive than negative responses were given, overall attitudes were less clear with regard to customizations to their own therapy style, options to modify the overall pacing of the game, repeat or change the number of game sessions, or modify characters. Therapists were also more neutral regarding the potential for young people themselves or the computer to make automatic customization.

Thematic Analysis

Thematic analysis of survey data and interviews identified the following 3 key themes: stages of use, impact on the delivery of therapy, and customization. As illustrated in Figure 3, 2 main themes contained multiple subthemes. For example, the stages of use theme included subthemes on training, deciding when and with whom to use Pesky gNATs, the approaches taken when introducing the game, and equally important, decisions regarding when to stop using the game.

Figure 3. Overview of the identified themes and subthemes.



Theme 1: Stages of Use

Training

The majority of therapists indicated that they completed the full online training. The training was valued and considered helpful in understanding the concepts and approaches to delivering therapy with Pesky gNATs. However, therapists also noted that the practical experience in using the game was important in gaining full value from the program:

> *My experience has been that becoming familiar with the material in advance and using the support available on the website means that I am in a position to anticipate and/or respond to any signs that a young person needs more time to explore ideas and their own feelings during a session.* [P20, survey]

A minority of therapists reported completing most or some of the training, of which one said that the game is easy to work with without the training. Several noted time constraints, with one explaining that they had insufficient time to complete training regarding the mobile phone app and therefore decided to focus on material relating to face-to-face sessions and supplement this with the paper workbook.

Deciding When and With Whom to Use Pesky gNATs

As per the recommended useage guidelines of Pesky gNATs, it was typically used to address depression and anxiety. Age, willingness to engage in face-to-face therapy, the complexity of the difficulties faced, and the intellectual abilities of the young people were key considerations in decisions to use the game. However, there were significant variations across therapists regarding the young people and situations where the game was deemed most appropriate. Some felt it was best suited for mild to moderate anxiety, but less fitting for severe cases. For example, a therapist explained that young people with severe psychological difficulties needed something simple as they might not have the patience or ability to engage with the game content. They continued that the game had been very effective in addressing low mood of children with regular ability levels. In contrast, another therapist described complex cases where Pesky gNATs was used after confirming the children's abilities and understanding of concepts needed for the game:

> *Some people on the autism spectrum and children with attachment disorder and borderline learning disabilities as well. So very complex children...I would do a preassessment of their actual understanding of thoughts, feelings, and behaviors...The other thing would be, have they got the intellectual ability to cope with the game and the patience to manage the game.* [P12, interview]

Importantly, therapists reflected that the decisions to use Pesky gNATs were not taken in isolation of other options. They described it as one of the tools they used to support the delivery of interventions. Although some therapists appeared to consider Pesky gNATs as a default option, others explained that the game was only introduced as an alternative when young people were reluctant to engage with the therapy. Overall, the therapists' decisions regarding the use of the game appeared to evolve with experience, with some describing decision-making steps:

> *...the first thing I would consider is, are they going to jell well with the material? Do they see the purpose of that material and what it's trying to say...[Secondly] any learning difficulties. The third consideration...have I got the time as an educational psychologist because our time is so limited?* [P4, interview]

One of the most surprising findings was that several therapists felt Pesky gNATs was not suitable for young people who regularly played computer games, as they might have different expectations from a game and miss the excitement of more action-oriented gameplay, described by one therapist as *killing things*.

Introducing Pesky gNATs to Young People

The video training emphasized that the manner in which Pesky gNATs is introduced is important. In particular, although we aimed to produce a high-quality game, Pesky gNATs cannot, nor is it intended to, compete with commercial games in terms of state-of-the-art graphics and the scale of the gameplay. The training suggested that therapists introduce the game to young people in a way that managed expectations and explained that Pesky gNATs was likely be a different experience to regular games. Participants confirmed the importance of this approach. One explained that they introduced the game by contrasting it to regular talk-based therapy. Another explained how an appropriate introduction was effectively used with a 17-year old, who might otherwise have felt too old for Pesky gNATs:

> *I used it with one pupil who was 17 years old because we almost looked at it with a bit of humor, almost tongue-in-cheek and alright, this is going to be quite corny and I was making a point. Giving them that little caveat, I think it was still helpful for her.* [P4, interview]

Therapists typically had traditional sessions with children before introducing Pesky gNATs. Some had 1 or 2 intake sessions before using the game, whereas others indicated it could take up to 3 or 4 sessions. One therapist described how it was useful to have a contract as part of the introduction and that this could help to maintain engagement:

> *I have made a couple of contracts with boys that I think could use it. Like, you do the first four sessions and then we'll talk about it and if you absolutely do not like it, we'll do something else and everybody's chosen to finish it. Some people have tolerated it, I should say, but they did get some things out of it. Using that contract with them really works.* [P7, interview]

Another therapist explained that it had been useful to introduce Pesky gNATs to parents by having them join the first session. This allowed parents to learn what happens in therapy and develop a better understanding of how to support their children.

Deciding to Discontinue Use

Pesky gNATs was not always helpful. When children were dissatisfied or when the therapists noticed that the game was not helping, it was discontinued. In some cases, children did not understand the CBT concepts, because of either comprehension issues or distraction. A therapist described a boy who switched to *gamer mode* and wanted to get to the fun and challenging parts of the game quickly. He had lost focus on the therapeutic concepts and had to be slowed down to focus, which had become a frustration in itself:

> *It was like he'd switched to gaming mode instead of CBT mode. So, for him it didn't help the therapeutic relationship in that session because it got in the way.* [P10, interview]

In contrast, some young people found the game boring or childish. For example, a therapist found that 16- and 17-year-old boys might feel *too cool* for the game. It is important to recognize these cases and where appropriate, discontinue its use.

Interestingly, the decision to stop using the game could be helpful to the overall therapeutic process. A therapist described a case where she introduced the game to a girl who was reluctant to engage in therapy. The girl also disliked the game and they discontinued its use. However, the therapist described how an improved relationship then emerged:

> *Even though we didn't use the game, I think it actually helped move the therapeutic relationship on, because we were able to have a conversation about the game.* [P10, interview]

Often it was a collaborative decision to stop using Pesky gNATs and try alternatives. A therapist described conversations with young people about why the game did not work and encouraged them to try it for at least two sessions. If the game remained unhelpful, the paper workbook with the same concepts was used instead, effectively leaving out the technology but keeping the narrative.

Theme 2: Impact on the Delivery of Therapy

Engagement and Changes in the Therapeutic Interaction

Most participants considered the game helpful and valued it as part of their practice. Therapists were somewhat surprised by how much young people learned and felt it was a *really helpful way* to introduce concepts that were ordinarily difficult to explain. Others described how the game and narrative helped with engagement, describing how navigating and exploring the island together was a good way to break the ice, build rapport, and maintain engagement:

> *They were excited about coming back to play Pesky gNATs, which people have not found that kind of excitement for them. The parents have not ever seen them excited to come back.* [P7, interview]

Changes in the traditional face-to-face interaction were identified as an important factor in improving engagement. Therapists described how the intensity that some young people experienced in face-to-face conversations was reduced:

> *I really like the idea of Pesky gNATs because you've got this tool to mediate that interaction and to build on. And I guess, to use as a prompt and an extension so the child's hearing the messages from a different place, as well as from you.* [P10, interview]

> *All three boys would have been reluctant to engage with paper-based activities without the motivation and context provided by the game. In each case, the program was helpful to lessen the intensity of one-to-one sessions. One young person particularly enjoyed some of the "goofy" aspects, such as the celebratory dance and the humor (intended or not) was very helpful to him.* [P35, survey]

As illustrated by this quote from P35, the adoption of a digital medium and elements of humor were also seen as important. However, although this attitude was common, it needs to be contrasted with other cases where some older children and gamers were alienated by the game-based approach and felt the humor was not *cool*.

Contrasting reactions were also evident in more specific functionality within Pesky gNATs. The use of standardized questionnaires provided a powerful example. Most therapists and children were happy with the in-game use of psychological scales to identify difficulties and visualize progress. However, in a minority of situations they were seen as detrimental. One therapist described how displaying questionnaires' results had the potential to put undesired labels on children:

> *I find it hard to then have a conversation around, "We're not looking at you being OCD, we're not thinking about a diagnosis. What we're thinking about is you having rigid and recurring thoughts and how we can address that", but it's hard to address something when you're given a label.* [P10, interview]

The mobile phone app was mostly perceived as useful for homework between sessions, but some therapists reported difficulties with children's engagement. There was too much competition for attention on phones, and some children did not own a phone, did not have access to their parents' device, or could only install apps with their parent's permission.

Usability and Interaction Issues

Several therapists noted that they encountered logistical difficulties in installing the software or loading and saving files on restricted organizational systems. When using the game, some felt that young people found the graphics and animations disappointing. Sometimes this was because of graphical glitches, such as a slow frame rate on older machines. Updating or upgrading the visual appearance of the game was suggested frequently when therapists were asked about potential improvements.

Age- and disorder-specific interaction difficulties were also noted. A therapist described how a 10-year old with anxiety difficulties could take a long time to complete text-based answers to questions in the game, feeling the answers had to be spelled and punctuated perfectly, which caused further anxiety. They suggested replacing some text input with multiple-choice answers or emoticons to simplify the interactions.

Some therapists suggested that the overall experience could be enhanced by adding more fun to parts of Pesky gNATs *to take a break if it feels too content heavy*. For example, having more mini-games as rewards after talking to characters and learning new concepts. In addition, therapists recommended additional gamification of therapeutic exercises:

> *More game type engagement could be very valuable. Games to identify and match gNATs to thoughts could be good. Further gamifying the therapeutic challenges could be very valuable and motivating.* [P22, survey]

The majority of therapists felt the structure of sessions and pacing of CBT concepts were appropriate. However, some did encounter difficulties with younger children's comprehension of concepts. Therapists noted that they would like the flexibility to go back and repeat previous levels. This is not currently possible; however, one therapist explained that using the paper workbook to revisit materials provided a workable alternative. Pacing was also considered from an interaction perspective. In addition to finding game characters, one therapist explained that children liked swimming and playing with beachballs and kicking them in the water and asked if there could be more similar interactions. However, as with other aspects of the game, there were variations in experiences. A minority felt that navigation on the island was too slow for some children and a quicker way to navigate would be an improvement.

Adaptation and Unexpected Use

In some cases, therapists supplemented the content of the game. One described how the homework exercises in Pesky gNATs did not match the immediate challenges of the child, and alternative homework was created that remained close to the game narrative:

> *She would get anxiety that there was always a burglar breaking into the house and so then, she didn't want to sleep in her own bed. So, we did homework around some classic exposure therapy and helping her parents. So, it didn't dovetail exactly with the game. Although, we did call it her Pesky gNat. We named it. It was this burglar gNat.* [P5, interview]

When working with young children it was not uncommon for therapists to involve parents or guardians. Although Pesky gNATs was not designed to explicitly support such involvement, several therapists did involve parents. For example, parents were invited to join the first session or were updated via email. A therapist explained that the homework exercises were useful to engage the family of a child. They felt that the program could be extended to more explicitly engage parents and educate them about the therapy sessions, so their support can be optimized. In contrast, others were more cautious, noting that some parents showed low engagement in the therapy of their children or that parents were part of the therapeutic difficulties. One explained that care should be taken when designing technology that requires children to engage their parents, as it might set up the child to fail if parents are resistant to this engagement.

Theme 3: Customization

Typically, therapists were open and positive to potentially customizing Pesky gNATs. Therapists would consider adjusting the game for individual children, as long as it was time-efficient, for example, repeating levels, adjusting in-game examples of thoughts and feelings, or difficulty of challenges and dialogues. Nevertheless, one therapist explained that they would not know how to make the game more engaging as they were not game developers. Another therapist worried that too much freedom could raise issues with the fidelity of the CBT process:

> *I don't think I would...because I know me well enough to know that I would skip over parts that are necessary, maybe to rush it a little bit and the one thing I can do now is say, "we've got to go from beginning to end. It's set up that way, I can't just do the fun things with you". That is a safeguard for me and the kids.* [P7, interview]

Some therapists liked the user-centered approach and favored children's ownership in therapy. They felt that allowing children to make changes to the gameplay and avatars could support engagement. Other therapists warned of potential distractions of such approaches because of the psychological difficulties that can arise:

> *I would just be concerned that children using it, particularly if they suffer with anxiety, need support, and to feel continued. Perhaps making changes together, if the child is a bit older?* [P8, survey]

Automatic customization in game mechanics by the computer were also considered. For example, mechanics that adjusted the difficulty of challenges based on the progress and performance of players. Some therapists were skeptical of automatic adjustments, fearing that it might impede professional judgements. Some also felt they had insufficient technical experience to appraise automatic adjustments.

Discussion

Principal Findings

Although our ongoing RCT [30] will provide more robust results on the clinical efficacy of Pesky gNATs, the naturalistic results reported here provide further evidence that serious games can support mental health interventions in real-world settings. The results indicate that the helpfulness of Pesky gNATs can be explained by factors including child-friendly, metaphor-driven narratives; playful elements; the fact that it is played alongside a therapist; and the changed therapeutic dynamics the game supports. Notably, we found evidence that serious games may be less appropriate for those who regularly play games (*gamers*).

Viewed more broadly, this deployment study also allowed us to identify generalizable recommendations relevant to the wider community, including researchers and designers seeking to maximize the impact of digital technologies in real-world settings. These include the following:

- Expect variability in use and support therapist autonomy
- New technologies become part of an overall toolbox
- Report implementation data, but also support a community of practice
- Carefully consider different approaches to customization and adaptability
- Deciding when and how to discontinue use is also important and can be helpful

Expect Variability in Use and Support Therapist Autonomy

Variability in use is perhaps the clearest message to emerge from our study. The young people for whom the therapists felt Pesky gNATs was most appropriate varied widely in age and psychological difficulties. There was also significant variability with regard to when therapists chose to introduce the game (eg, as a default or only with reluctant young people).

Variability is likely to reflect the reality of technology use in the real-world and as such it should be anticipated and where possible supported. Recognizing the role of the therapist as a decision-maker and considering how this autonomy can be supported is an important design consideration. Pesky gNATs was specifically intended to support therapists' existing practices and recognize the value of therapeutic relationships. The game does not seek to replace therapists; rather it is aimed at subtly changing the dynamics of face-to-face interventions and introducing CBT concepts in a child-friendly manner, while ensuring that the therapist retains control and guides the overall intervention.

It is important to note that in this context, variable use does not indicate suboptimal use. Equally, we do not advocate inconsistent or inappropriate use. For example, access to Pesky gNATs is subject to manual screening of therapists' experience and qualifications. Although this process is time-consuming and subject to abuse, we believe it is essential. Furthermore, although therapists in our study did not agree on the situations in which Pesky gNATs was most appropriate, they described careful individual decision-making processes regarding its use. The ability to understand the CBT concepts in the game and regular, reliable meeting schedules played an important role in such decisions. The online training for Pesky gNATs also provided guidance on such decisions. More broadly, researchers should accept the ongoing responsibility to gather data that helps to establish the boundaries of acceptable use and update and evolve guidelines accordingly.

New Technologies Become Part of an Overall Toolbox

In discussing common misconceptions about mental health technologies, Mohr et al argue that "research literature typically describes and evaluates mental health technologies as if they were products" [32]. Rather than viewing new technologies as products that can mediate change in and of themselves, they argue it is more appropriate to view them as technology-enabled services, which serve to support the overall service or therapeutic process. Supporting the autonomy of therapists will be an important factor in advancing this new perspective. However, it is also important to recognize that individual programs or services will not exist in isolation. Rather, they will become part of the broader toolbox from which therapists will choose.

This deployment study clearly showed that Pesky gNATs was not used in isolation. Rather, therapists considered it as one option among many. For some it was a preferred option, whereas others only used it when young people were resistant to more traditional face-to-face approaches. When it was used, therapists often used it in combination with other approaches. Sometimes these alternatives directly mirrored activities in Pesky gNATs. For example, some young people preferred to do mindfulness exercises directly with their therapist, instead of using the game. On other occasions, therapists supplemented Pesky gNATs with activities entirely separate to the program, based on the emerging needs of young people (eg, guided exposure). Providing a paper workbook as an alternative or backup to the game and mobile phone app was also an important part of the therapist toolbox.

Report Implementation Data, but Also Support a Community of Practice

As noted in the Introduction, there is growing recognition that implementation data should be routinely reported [32,34]. We support such recommendations and believe that shared deployment data have clear benefits for the research community. In this context, it is critically important that both the successes and failings of systems are shared. It is also important to report how technologies fit with the broader ecosystem, which is multifaceted and involves overlapping technical, organizational, and social factors. This is illustrated by the difficulties some therapists experienced when installing Pesky gNATs on organizational computers. In addition, some young people did not have their own smartphones or were dependent on parental permission to install the app. Sharing such experiences will enable us to develop generalizable theories and design principles that support more impactful technologies, for example, guidelines for customization or supporting therapist autonomy.

Beyond the research community, we also recommend a further use of deployment data—supporting the creation of communities of practice. Therapists in our study clearly valued the online training. However, they explained that the experience of using the game was critical. Although each therapist needs to gain this experience first-hand, there is a strong argument for supporting online communities where experiences of using new technologies can be shared. Specific examples with regard to Pesky gNATs include techniques for introducing the game and managing expectations, or the use of therapeutic contracts. The creation of communities of practice may also be effective in encouraging early adopters of digital technologies to become active, participatory co-designers in working with developers and researchers to improve interventions.

Carefully Consider Different Approaches to Customization and Adaptability

Our study explored 3 distinct approaches to customization. Across all approaches, therapists clearly recognized the need for systems to maintain fidelity of a core therapeutic model. The following are the customization approaches:

- *Customization by therapists* was generally viewed positively, although an important distinction can be drawn between changes to therapeutic content versus changes to interaction-specific features. For example, the option of adjusting the in-game examples of thoughts and feelings or content of in-game dialogues was viewed favorably. Therapists were less willing or confident in their ability to modify game-based or technical interaction features.
- *Customization by young people* was viewed less positively, even by therapists who favored young peoples' ownership of therapy. However, specific customizations, such as avatar personalization, were viewed positively. Indeed, recent evidence has suggested that avatar personalization can improve overall engagement [49]; thus, there may be significant value in enabling this form of customization.
- Intelligent *customization by the computer* was viewed favorably by some, but drew skepticism from many therapists, partially because of concerns over a loss in control of the process and also a lack of clear understanding of what intelligent customization would entail. If intelligent customization is to become a feature in mental health technologies, clarity around the targeted adaptation and factors such as transparent and explainable Artificial Intelligence will be critical.

Deciding When and How to Discontinue Use Is Also Important and Can Be Helpful

Adherence to study protocols is a defining characteristic of RCTs. Although many RCTs seek to understand and follow up with dropouts, completion of the full intervention, including postintervention data collection, is essential to establish efficacy. In real-world contexts, and when individual technologies are considered as part of a toolbox, the decision to stop using technologies can be viewed more flexibly and with less focus on adherence. If a technology alienates a young person to the extent that they withdraw from therapy, this would be a significant cause for concern. However, evidence from this study suggested that therapists and young people often made joint decisions to stop using the game. In such cases, discussions about why the young person disliked the game—and what alternative approaches might be useful—were often helpful in and of themselves.

This more flexible view of technology—where systems are engaged for as long as they are useful and ongoing use is based on users' emerging needs and reactions—aligns with arguments previously made by Doherty et al [50] in the context of online CBT. More recently, Lederman et al [51] have explored concordance as a valuable perspective through which to view the design of technologies. Here, the overall effectiveness of treatment, support for core treatment goals, and collaborative decision-making involving therapist and clients, again take precedence over strict adherence. The evidence in this deployment study supports Lederman's argument that concordance is a valuable lens through which to understand the effectiveness of new mental health technologies.

Limitations

We acknowledge the small sample size and participants' self-selection bias, as only therapists who had decided to use the program were invited to participate. This study was not designed to provide robust evidence regarding efficacy; no standardized data regarding clinical outcomes were collected. Furthermore, therapists used Pesky gNATs with diverse groups and in diverse contexts and this may have influenced the specific perspectives of therapists.

Conclusions and Future Work

This study has provided further evidence that serious games can support mental health interventions for young people in real-world settings. More broadly, it provided evidence for the potential of naturalistic deployment studies to complement traditional approaches to evaluation, for example, RCTs. Deployment studies provide a valuable means of understanding how new technologies become part of an overall therapeutic toolbox and help in the development of technologies that are more easily integrated to the mental health ecosystem. Variability in use should be expected in real-world settings. Effective training, support for therapist autonomy, careful consideration of different approaches to customization, the reporting of deployment data, and support for shared communities of practice can play an important role in supporting variable, but effective, use. Future research could include larger-scale deployment studies that bridge the gap between controlled studies and studies in naturalistic settings.

Acknowledgments

This research is part of TEAM, an Innovative Training Network funded by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement number 722561.

Conflicts of Interest

GOR and DC are the creators of Pesky gNATs, which is owned by *Handaxe Community Interest Company (CIC), a not-for-profit company*. GOR and DC are directors of Handaxe CIC but receive no financial reward for this role.

Multimedia Appendix 1

The structure of game levels and cognitive behavioral therapy concepts in Pesky gNATs.

[PDF File (Adobe PDF File), 42KB - mental_v6i5e12430_app1.pdf]

References

1. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
2. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
3. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
4. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
5. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
6. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, et al. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
7. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
8. Great Britain Department of Health and Social Care. 2011. No health without mental health: a cross-government mental health outcomes strategy for people of all ages URL: https://www.gov.uk/government/publications/no-health-without-mental-health-a-cross-government-outcomes-strategy [accessed 2019-03-01] [WebCite Cache ID 778gZxjHc]
9. World Health Organization. 2013. Comprehensive mental health action plan 2013-2020 URL: https://www.who.int/mental_health/publications/action_plan/en/ [accessed 2019-03-01] [WebCite Cache ID 778hLeXvN]
10. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
11. Buckley S, Cannon M, Chambers D, Coughlan H, Duffy M, Gavin B, et al. International Association of Youth Mental Health. 2011. The International Declaration on Youth Mental Health URL: http://publichealthwell.ie/node/114992 [accessed 2019-03-01] [WebCite Cache ID 77NYhsmfM]

12. Mental Health Partnerships. 2015. Future in mind: Promoting, protecting and improving our children and young people's mental health and wellbeing URL: https://mentalhealthpartnerships.com/resource/future-in-mind/ [accessed 2019-03-01] [WebCite Cache ID 778ikPeuo]
13. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
14. Gore FM, Bloem PJN, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: a systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
15. Copeland WE, Adair CE, Smetanin P, Stiff D, Briante C, Colman I, et al. Diagnostic transitions from childhood to adolescence to early adulthood. J Child Psychol Psychiatry 2013 Jul;54(7):791-799 [FREE Full text] [doi: 10.1111/jcpp.12062] [Medline: 23451804]
16. O'Reilly G, Coyle D. Pesky gNATs. Bristol; 2015. URL: http://peskygnats.com/ [accessed 2018-09-24] [WebCite Cache ID 72fvWgHiT]
17. Khazaal Y, Favrod J, Sort A, Borgeat F, Bouchard S. Editorial: computers and games for mental health and well-being. Front Psychiatry 2018;9:141 [FREE Full text] [doi: 10.3389/fpsyt.2018.00141] [Medline: 29706906]
18. Shah A, Kraemer KR, Won CR, Black S, Hasenbein W. Developing digital intervention games for mental disorders: a review. Games Health J 2018 Aug;7(4):213-224. [doi: 10.1089/g4h.2017.0150] [Medline: 30106642]
19. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
20. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
21. Lucassen M, Merry S, Hatcher S, Frampton C. Rainbow SPARX: A novel approach to addressing depression in sexual minority youth. Cogn Behav Pract 2015;22(2):203-2016. [doi: 10.1016/j.cbpra.2013.12.008]
22. Cooney P, Jackman C, Coyle D, O'Reilly G. Computerised cognitive-behavioural therapy for adults with intellectual disability: randomised controlled trial. Br J Psychiatry 2017 Aug;211(2):95-102. [doi: 10.1192/bjp.bp.117.198630] [Medline: 28596245]
23. Tak YR, van Zundert RM, Kuijpers RC, van Vlokhoven BS, Rensink HF, Engels RC. A randomized controlled trial testing the effectiveness of a universal school-based depression prevention program 'Op Volle Kracht' in the Netherlands. BMC Public Health 2012 Jan 10;12:21 [FREE Full text] [doi: 10.1186/1471-2458-12-21] [Medline: 22233510]
24. Coyle D, Matthews M, Sharry J, Nisbet A, Doherty G. Personal Investigator: a therapeutic 3D game for adolecscent psychotherapy. Interactive Tech Smart Ed 2005 May 31;2(2):73-88. [doi: 10.1108/17415650580000034]
25. Brezinka V. Ricky And The Spider - a video game to support cognitive behavioural treatment of children with obsessive-compulsive disorder. Clin Neuropsychiatry 2013;10(3). [doi: 10.5167/uzh-93917]
26. Brezinka V. Treasure Hunt - a serious game to support psychotherapeutic treatment of children. Stud Health Technol Inform 2008;136:71-76. [doi: 10.3233/978-1-58603-864-9-71] [Medline: 18487710]
27. Stasiak K, Hatcher S, Frampton C, Merry SN. A pilot double blind randomized placebo controlled trial of a prototype computer-based cognitive behavioural therapy program for adolescents with symptoms of depression. Behav Cogn Psychother 2014 Jul;42(4):385-401. [doi: 10.1017/S1352465812001087] [Medline: 23253641]
28. Coyle D, McGlade N, Doherty G, O'Reilly G. Exploratory evaluations of a computer game supporting cognitive behavioural therapy for adolescents. New York, NY, USA: ACM; 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC, Canada p. 2937-2946 URL: https://doi.org/10.1145/1978942.1979378
29. Rebello TJ, Marques A, Gureje O, Pike KM. Innovative strategies for closing the mental health treatment gap globally. Curr Opin Psychiatry 2014 Jul;27(4):308-314. [doi: 10.1097/YCO.0000000000000068] [Medline: 24840160]
30. McCashin D, O'Reilly G, Coyle D. ISRCTN. 2019. Is the computer game Pesky gNATs an effective intervention forchildren experiencing low mood or anxiety? URL: http://www.isrctn.com/ISRCTN60159987 [accessed 2019-04-17] [WebCite Cache ID 77i6DIHhX]
31. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. New York, NY, USA: ACM; 2011 Presented at: SIGCHI conference on human factors in computing systems; May 7-12, 2011; Vancouver, BC, Canada p. 3063-3072.
32. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research... and three things we can do about them. Psychiatr Serv 2017 May 1;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
33. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]

34. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 6;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
35. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
36. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
37. Twomey C, O'Reilly G. Effectiveness of a freely available computerised cognitive behavioural therapy programme (MoodGYM) for depression: meta-analysis. Aust N Z J Psychiatry 2017 Mar;51(3):260-269. [doi: 10.1177/0004867416656258] [Medline: 27384752]
38. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2015;351:h5627 [FREE Full text] [Medline: 26559241]
39. Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RC, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: a school-based program (Op Volle Kracht) and a computerized program (SPARX). Behav Res Ther 2016 May;80:33-42. [doi: 10.1016/j.brat.2016.03.005] [Medline: 27019280]
40. Wijnhoven L, Creemers D, Vermulst A, Scholte R, Engels R. Randomized controlled trial testing the effectiveness of a depression prevention program ('Op Volle Kracht') among adolescent girls with elevated depressive symptoms. J Abnorm Child Psychol 2014 Feb;42(2):217-228. [doi: 10.1007/s10802-013-9773-5] [Medline: 23893066]
41. O'Reilly G. Pesky gNATs! Using computer games and smartphone apps to teach complex cognitive behavioural therapy and mindfulness concepts to children with mental health difficulties. In: The Use of Technology in Teaching and Learning. Washington: Society for the Teaching of Psychology; 2018:231-247.
42. Bandura A, Walters R. Social Learning Theory. Englewood Cliffs, NJ: Prentice-hall; 1977.
43. Beck J. Cognitive Behavior Therapy: Basics and Beyond. New York: Guilford press; 2011.
44. Dunn J. The Beginnings of Social Understanding. Boston: Harvard University Press; 1988.
45. Flavell J. The developmental psychology of Jean Piaget. In: The University Series in Psychology. Princeton, NJ: Van Nostrand; 1963.
46. Vygotsky L, Cole M. Mind in Society. Boston: Harvard University Press; 1978.
47. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
48. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. London: Sage; 2013.
49. Birk MV, Mandryk RL. Improving the efficacy of cognitive training for digital mental health interventions through avatar customization: crowdsourced quasi-experimental study. J Med Internet Res 2019 Jan 8;21(1):e10133. [doi: 10.2196/10133] [Medline: 30622095]
50. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM; 2012 Presented at: CHI'12; May 5-10, 2011; Austin, Texas p. 1421-1430.
51. Lederman R, Gleeson J, Wadley G, D’alfonso S, Rice S, Santesteban-Echarri O, et al. Support for carers of young people with mental illness: design and trial of a technology-mediated therapy. ACM Trans Comput Hum Interact 2019 Feb 23;26(1):1-33. [doi: 10.1145/3301421]

Abbreviations

3D: 3-dimensional
CBT: cognitive behavioral therapy
RCT: randomized controlled trial

Edited by M Birk; submitted 05.10.18; peer-reviewed by M Tielman, RJ Beun, A Mahnke; comments to author 13.01.19; revised version received 21.03.19; accepted 04.04.19; published 09.05.19.

Please cite as:
van der Meulen H, McCashin D, O'Reilly G, Coyle D
Using Computer Games to Support Mental Health Interventions: Naturalistic Deployment Study
JMIR Ment Health 2019;6(5):e12430
URL: https://mental.jmir.org/2019/5/e12430/
doi:10.2196/12430
PMID:31094346

Original Paper

Exploring User Needs and Preferences for Mobile Apps for Sleep Disturbance: Mixed Methods Study

Melissa Aji[1,2], BPsych (Hons); Christopher Gordon[2,3,4], PhD; Dorian Peters[5], BHA, MMDes; Delwyn Bartlett[3,6], PhD; Rafael A Calvo[5,7], PhD; Khushnood Naqshbandi[5], BIT (Hons); Nick Glozier[1], PhD, MBBS, FRANZCP

[1]Brain and Mind Centre, Central Clinical School, Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Camperdown NSW, Australia

[2]CRC for Alertness, Safety and Productivity, Melbourne, Victoria, Australia

[3]Centre for Sleep and Chronobiology, Woolcock Institute of Medical Research, Glebe, Australia

[4]Susan Wakil School of Nursing and Midwifery, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[5]School of Electrical and Information Engineering, University of Sydney, Sydney, Australia

[6]Central Clinical School, Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[7]Dyson School of Design Engineering, Imperial College London, London, United Kingdom

Corresponding Author:
Nick Glozier, PhD, MBBS, FRANZCP
Brain and Mind Centre
Central Clinical School, Sydney Medical School
Faculty of Medicine and Health, University of Sydney
Level 5, Professor Marie Bashir Centre, Missenden Road
Camperdown NSW, 2050
Australia
Phone: 61 2 9515 1596
Email: nick.glozier@sydney.edu.au

Abstract

Background: Mobile health (mHealth) apps demonstrate promise for improving sleep at scale. End-user engagement is a prerequisite for sustained use and effectiveness.

Objective: We assessed the needs and preferences of those with poor sleep and insomnia to inform the development of an engaging sleep app.

Methods: We triangulated results from qualitative (focus groups and app reviews) and quantitative (online survey) approaches. A total of 2 focus groups were conducted (N=9). An online survey tested themes identified from the focus groups against a larger population (N=167). In addition, we analyzed 434 user reviews of 6 mobile apps available on app stores.

Results: Common focus group themes included the need to account for diverse sleep phenotypes with an adaptive and tailored program, key app features (alarms and sleep diaries), the complex yet condescending nature of existing resources, providing rationale for information requested, and cost as a motivator. Most survey participants (156/167, 93%) would try an evidence-based sleep app. The most important app features reported were sleep diaries (148/167, 88%), sharing sleep data with a doctor (116/167, 70%), and lifestyle tracking (107/167, 64%). App reviews highlighted the alarm as the most salient app feature (43/122, 35%) and data synchronization with a wearable device (WD) as the most commonly mentioned functionality (40/135, 30%).

Conclusions: This co-design process involving end users through 3 methods consistently highlighted sleep tracking (through a diary and WD), alarms, and personalization as vital for engagement, although their implementation was commonly criticized in review. Engagement is negatively affected by poorly designed features, bugs, and didactic information which must be addressed. Other needs depend upon the type of user, for example, those with severe insomnia.

(JMIR Ment Health 2019;6(5):e13895) doi:10.2196/13895

KEYWORDS

mobile apps; mHealth; sleep

Introduction

Sleep disturbances are widespread, impacting 33% to 45% of adults [1]. The most common sleep disorder is insomnia which is underrecognized and undertreated [2]. The most limiting treating factor is a shortage of treatment resources including trained clinicians [3]. In response, there has been increasing interest in harnessing mobile technology for sleep health. The mobile app marketplace has expanded rapidly with approximately 325,000 mobile health (mHealth) apps now available across major app stores [4].

Existing mHealth sleep apps offer a broad range of functions including sleep tracking, education, alarm clocks, and sound recording during sleep. A recent review examining mobile phone interventions for sleep suggested that mobile apps are effective in improving symptoms and sleep quality [5]. However, engagement is often poor: Approximately half of respondents who download health apps no longer use these owing to poor user experience [6]. Moreover, users are likely to abandon an app that fails to provide immediate engagement [7].

Critical to developing an engaging tool is an understanding of the needs, contexts, and preferences of end users, including functional (ie, what people need and expect an app to do) and nonfunctional requirements (ie, use of relevant and appealing language and visuals). We also need to understand the context of technology use (eg, will users sleep with their phone by the bed and use other sleep aids with the app?) and the characteristics of prototypical users (personas). Therefore, in developing a new sleep app [7], we undertook a mixed methods user-centric approach to determine the needs and preferences for sleep apps in those with poor sleep or insomnia.

Methods

Design

We triangulated data [8] from 3 end user–centric methods: focus groups, an online survey, and an analysis of app reviews to give a holistic understanding by harnessing both qualitative research to provide rich insights and quantitative research to examine trends in a larger population. This counterbalances inherent biases in each method with strengths inherent to the other methods.

We addressed poor quality sleep as a continuum from disturbed sleep to insomnia for the focus groups and the survey. Both sample participants were aged ≥18 years and screened for poor sleep and insomnia using the Pittsburgh Sleep Quality Index (PSQI) and Insomnia Severity Index (ISI), respectively.

Ethical approval for the focus groups and survey was obtained from the Sydney Local Health District Ethics Committee (X17-0177). Before participation, focus group participants provided written consent and survey participants provided digital consent.

Pittsburgh Sleep Quality Index

The PSQI [9] assesses subjective sleep quality over the previous month. It comprises 7 components added to yield one *global* score (range 0-21). Higher scores represent poorer subjective sleep quality. A score of above 5 indicates poor sleep quality.

Insomnia Severity Index

The ISI [10,11] is a 7-item questionnaire (score range 0-28) assessing insomnia symptom severity over the past 2 weeks with higher scores indicating greater insomnia symptom severity and moderate-severe insomnia defined as >15.

Focus Groups

Participants and Recruitment

Invitations were sent to participants on the Woolcock Institute of Medical Research database. Eligible participants were fluent in English and assessed for Insomnia Disorder (Diagnostic and Statistical Manual of Mental Disorders) by (1) a telephone screening checklist with a study coordinator, (2) score ISI > 15, and (3) PSQI ≥5 plus component 5 (sleep disturbance) total score of 1 or 2.

Procedure

In total, 2 activity-based workshops (1.5 hours each) were conducted by a user experience specialist (DP) using semistructured topic guides and activities. Workshops explored participants (1) experience of poor sleep, (2) experience using technology to help with sleep, and (3) ideas and preferences for a sleep app. Activities including collaborative ideation and paper prototyping explored several user experience questions drawn from literature on participatory design [12-14]. Participants generated ideas for desirable functionalities and app characteristics and provided feedback on draft screen designs for a prototype app designed by the research team. The data collected included participant-generated artifacts (ie, filled-in advertisement templates and screen design sketches), collections of feature ideas, field notes, and audio recordings and workshop transcripts.

Data Analysis

Participatory workshops were recorded and transcribed verbatim using NVivo (QSR International Pty Ltd Version 11.4.3, 2017). These transcripts were analyzed using thematic analysis [15], consistent with methods of analysis of generative participatory data [12].

Thematic analysis focused on categorizing data to inform development (1) features, content, and characteristics of an ideal app and (2) experience of sleep problems that could be addressed via technology. Initial coding (DP and KN) involved attaching labels to text segments to identify themes related to the research questions. Analysis progressed iteratively with independent rereading of transcripts and reexamining themes against the raw data to further refine the themes and identify subthemes. Discrepancies in coding were regularly discussed and resolved. A thematic schema was developed using FreeMind software 0.9.0.

Survey

Participants and Recruitment

An anonymous online survey was conducted between May and July 2018. Facebook and Instagram advertisements invited participants to participate in a 20-min survey to help design a mobile app for sleep. Individuals were eligible if online screening indicated positive for poor sleep (PSQI ≥5) or insomnia symptoms (ISI >15) and provided informed online consent. Upon completion, participants could enter a draw to win a gift card (valued at AUD $50). All participants could receive a summary of the findings by including their email address. Responses were stored on a password-protected database on a secure server.

Survey Content

The survey assessed (1) demographics, (2) mobile phone usage, (3) wearable device (WD) usage, (4) sleep environment, (5) preferences for a sleep app, and (6) previous health app and website usage.

In total, there were 34 questions, with an open-ended question about likes and dislikes of existing digital health technologies (Multimedia Appendix 1 for survey).

Data Analysis

Descriptive statistics were performed for all items. Chi-square tests compared proportions of sleep app preferences across groups defined by insomnia severity (ISI <17 vs ISI ≥17) and WD ownership (WD vs no WD). Analyses were performed using SPSS version 24.

App Reviews

Existing studies exploring user experience of online interventions often focus on perspectives of treatment completers or active engagers [16,17], limiting the sample to satisfied users who persist with the program. On the contrary, app reviews garner perspectives from a larger and more diverse sample. They are often written by users with varying satisfaction including those who are extremely dissatisfied and have given up on the app.

Sample

In February 2018, an electronic search was conducted on Google Play and the App Store (United States) using the following search terms: *sleep*, *sleep health*, and *sleep therapy*. Apps were included if they (1) were consumer targeted, (2) targeted adults, (3) included the term *sleep* in the app title or description, (4) included personalized sleep feedback, (5) allowed for sleep tracking with a WD, and (6) had an English interface.

Apps were excluded if they were developed with the sole function of sleep tracking, relaxation sounds, hypnosis, meditation, recording of snoring and sleep talking, and/or alarm clocks. Apps were also excluded if they (1) required health care provider guidance, (2) were only accessible via an employer, (3) had fewer than 10 reviews, or (4) were a companion app for a sleep measurement device.

Analysis of User Reviews

Qualitative content analysis was used to code the app user reviews [18]. A pilot coding scheme of major and subcategories was iteratively developed from a random sample of 165 reviews from the 6 apps (listed in the results). This coding scheme also adapted themes identified in previous studies reviewing app user feedback [19,20] including the Components of User Experience model [21]. In total, 2 coders (MA and KN) agreed on the final coding scheme (Multimedia Appendix 2).

The team independently coded user reviews line-by-line and classified as *praise* or *critique*. Coding disputes were resolved, and any new codes detected from analysis were discussed and incorporated into the coding scheme.

Results

Focus Groups

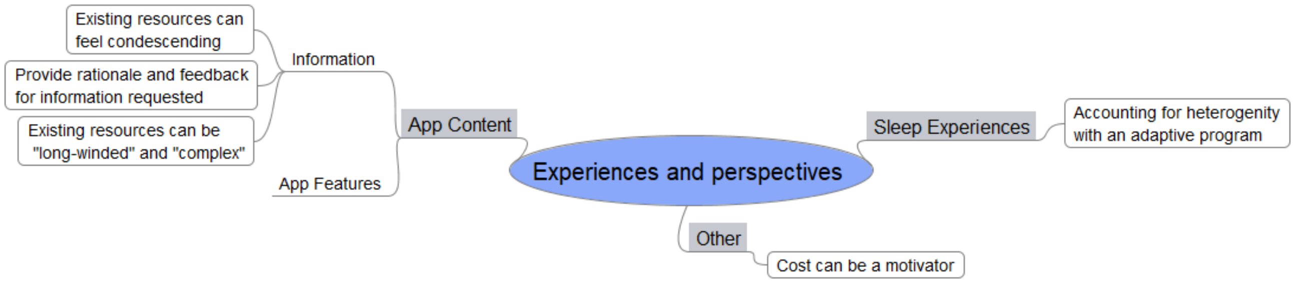
Sample

A total of 9 participants (6 males and 3 females, aged 21 to 70 years) were recruited.

Workshops

The participant’s needs and preferences for an engaging sleep app formed overarching themes of *sleep experiences*, *app content*, and *other* (Figure 1). A complete list is provided in Multimedia Appendix 3.

Figure 1. Thematic schema.



Main Themes and Insights

Sleep Experiences

Accounting for Heterogeneous Experiences With an Adaptive Program

Participants’ experience of sleep difficulty was diverse and included frequent night awakenings, trouble falling asleep, nightmares, snoring, partner disturbance, daytime stress and sleepiness, anxiety over not sleeping, and lack of a sleep rhythm. As such an ideal sleep program must be both customizable and adaptive to remain relevant to this wider population. Addressing the user’s specific sleep problems while catering for different lifestyles and schedules is crucial. A male participant’s statement illustrates both:

> *I find I'm so erratic with this...so a program that says, "Right, go to bed now"...sometimes I might go to sleep in a few minutes and have, not a good sleep, but doze. Other times, I can be as tired as possible, and I just can't sleep...So, you know, "go to sleep for five hours", and five hours later I'm still there looking at the ceiling. So if that didn't work, "let's try this approach".* [P3, FG2]

App Content

App Features

Participants generated feature ideas including:

- Alarms for waking up and staying awake when tired, challenging alarm turn-off sequences to prevent oversleeping and a selection of alarm sounds
- Audio features (meditation, podcasts, and relaxing sounds)
- Encouragement (signs of progress and rewards)
- Frequently Asked Questions were preferred to chatbots as the latter requires keyboard input
- Recommendations for helpful apps and websites
- Social features, in particular, the ability to share information among a community of poor sleepers. There was interest in tips from others who have successfully completed the program.

Unsurprisingly, there was a strong interest in sleep tracking and data visualization. As one participant [P3, FG2] stated, "Graphs appeal to me...Graphs, I like graphs." Participants expressed interest in objective data (eg, "How much you sleep per night, and how much is REM, and how much is deep sleep"[P3, FG1]) but also stressed the importance of subjective measures such as sleep *quality* or *satisfaction* explaining that duration alone can be misleading:

> *It says how many hours of sleep do you get a night, well, nowhere in there does it ask for the quality of sleep...If it's four, there are four hours, but it might be worth one out of ten.* [P2, FG2]

Tracking daily behaviors such as exercise or usage of treatments, for example, sleeping pills against sleep was important:

> *I think a graph that can be potentially smart enough to start linking some of these—you might try a tag, you know, "took magnesium",...and then you can graph your sleep, your quality of sleep against those tags...I like the tags.* [P3, FG2]

Information

Existing Resources Can Feel Condescending

Some participants found certain online sleep programs presented overly basic content and quizzes:

> *One of the really aggravating things about the program was...you had to complete this course in the week and they gave you these little things to do. And you were treated as an absolute moron, all the way through this thing. It was, you know, "now, if you exercise, you know, an hour before going to sleep, you get your adrenaline running and everything, so it's bad to exercise an hour before going to sleep". And so they told you all of these things, and then at the end of it, you had to do this quiz. Now, "Jenny has just exercised an hour before going to sleep. Is this good or bad for her?" And so you had to fill out these things, and you're just thinking, really? Do I really have to do this?...and you were just treated as an absolute moron. "Oh, you've got a hundred percent, well done." And you're just thinking, oh, my God, I've paid two hundred dollars to do this. And it was just really demeaning.* [P4, FG1]

Existing Resources Can Be Long-Winded and Complex

Several participants expressed frustration with existing online content and exercises for sleep being *too complicated*, *overwhelming*, *tedious*, and *long-winded*. One male participant explains:

> *But there's so much stuff. And all these suggestions and all these relaxation exercises and so much stuff, I just, I can't go through all this. It's just driving me nuts. They're just too complicated.* [P2, FG2]

This reflects the prevailing tension between the need for resources that are neither oversimplified and condescending nor too complex and overwhelming.

Provide Rationale and Feedback for Information Requested

Participants expressed frustration in filling out lengthy questionnaires without a clear rationale or immediate feedback on the data collected. A female participant explained:

> *Can I just say, with the questionnaires, the burden of answering so many questionnaires at times can be really—and I know that you've got to do this...but with an app, if it's explained at the outset, "we're going to ask you a lot of questions, and we apologise...for being such a difficult thing; but we're going to give you feedback straightaway on what we can see so far, or what the issues that might actually be identifiable at the outset". And if you actually had that right from the beginning...it would be much more intuitive, much more easy to use, and you'd actually feel benefit right from the start.* [P3, FG1]

Other

Cost Can Be a Motivator

Making a sleep app free may undermine long-term commitment. The female participant [P4, FG1] previously quoted as expressing frustration over the condescending program explained that she completed the program, but only as she had already invested money in it:

> *I just didn't think it was two hundred dollars well spent. And really the only thing they told me was "limit your sleep", and I didn't think that it worked for me.* [Participant]
>
> *But you made it to the end, so I'm curious, what do you reckon was the most motivating thing that got you to keep going even though you were...* [Facilitator]
>
> *The two hundred dollars...It's two hundred dollars and I'm a desperate woman...* [Participant]

Survey

Response Rates

The survey received 546 hits, with 186 people completing screening questionnaires. Of these, 19 were ineligible on the ISI and PSQI leaving 167 participants.

Demographics

Participants were aged between 18 and 79 years (mean 38.8, SD 11.9) and predominately female (88%; Table 1).

Technology Usage

Most participants (162/167, 97%) reported using a mobile phone daily (Table 2). A large proportion of participants reported almost always keeping their phone by their bed at night (142/167, 85%) and using their phone as an alarm clock (113/167, 68%). Approximately one-third (54/167, 32%) reported using a WD, with Fitbit (48%) as the most popular. The 3 most frequently used features of WD's were tracking for steps (48/54, 88%), heart rate (39/54, 72%), and sleep (35/54, 65%). Over half (94/167, 56%) reported using a health-related app. Among the 56% (93/167) using mobile apps for sleep, the most frequently used were meditation apps (45/93, 48%).

No differences were found in sociodemographic, mobile phone or WD use between those with high or low levels of insomnia symptom severity nor were there any sociodemographic differences identified between those who use a WD (all P>.05).

Table 1. Sociodemographic characteristics of respondents (n=167).

Variable	n (%)
Gender	
Female	147 (88.0)
Male	18 (10.8)
Prefer not to say	2 (1.2)
Age (years)	
18-30	48 (28.7)
31-40	55 (32.9)
41-50	33 (19.7)
51+	31 (18.6)
Education	
Secondary school	46 (27.5)
Diploma	25 (15.0)
Trade certificate	29 (17.4)
Bachelor's degree	45 (26.9)
Postgraduate degree	22 (13.2)
Employment[a]	
Full-time	53 (31.7)
Part-time	58 (34.7)
Student	17 (10.2)
Unemployed	20 (12.0)
Retired	11 (6.6)
Other	16 (9.6)

[a] Percentages do not add up to 100% as respondents were allowed multiple responses.

Table 2. Mobile phone, wearable device, and app usage (N=167).

Variable	n (%)
Mobile phone brand	
Samsung	73 (43.7)
Apple	70 (41.9)
Other	24 (14.4)
Wearable device	
Usage	
Yes	54 (32.3)
No	113 (67.7)
Brand	
Fitbit	26 (48.1)
Apple	10 (18.5)
Other	18 (33.4)
Purpose[a]	
Fitness	45 (83.3)
Health	36 (66.7)
Communication	23 (42.6)
Features used[a]	
Step tracking	48 (88.9)
Heart rate monitoring	39 (72.2)
Sleep tracking	35 (64.8)
Health app usage	
Yes	94 (56.2)
No	73 (43.7)

[a] Percentages do not add up to 100% as respondents were allowed multiple responses.

Sleep App Preferences

In total, 93% (156/167) of individuals indicated willingness to try an app proven to improve sleep. A sleep diary and tracking (148/167, 88%) the ability to share sleep data with their doctor (116/167, 70%), offline functionality (113/167, 68%), and tracking of lifestyle factors (107/167, 64%) were listed (Table 3). Only those participants with increased insomnia severity wanted to share sleep data with their doctor.

App Reviews

Descriptive Analysis

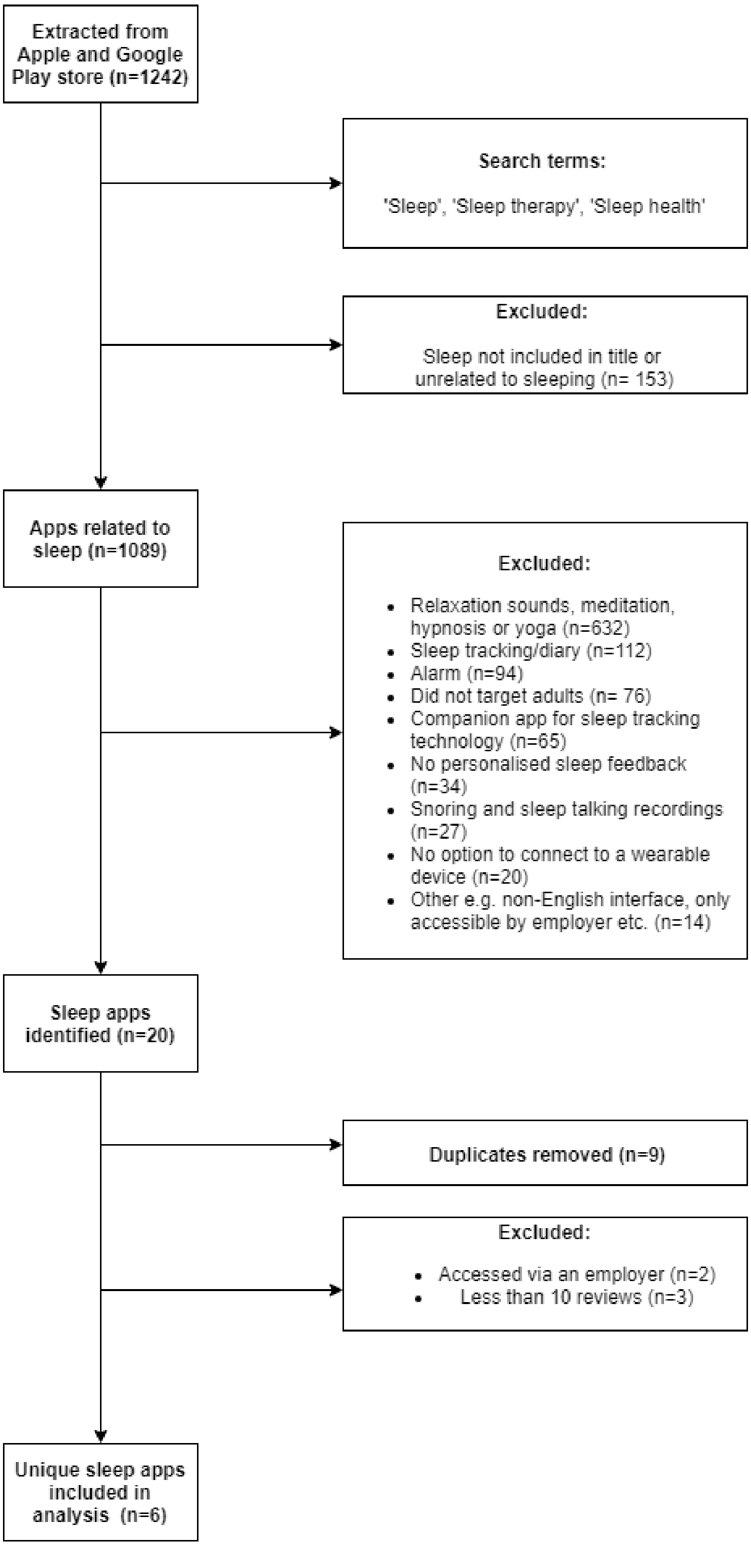
The initial search of all stores yielded 1242 apps, of which 1089 were sleep related (Figure 2), and 20 met the inclusion criteria. Apps were excluded if they were duplicates, had restrictive access, or <10 reviews. In total, 6 apps remained (Lark, Pillow, Sleep as Android, SleepHealth, Sleepio, and Sleeprate) and their reviews were analyzed (Multimedia Appendix 4 for app content). A total of 2 were in both in Google Play and the App Store, with the others found solely from either the App Store (3/6) or Google Play (1/6). Reviews for each app ranged from 11 and 246,240 (mean 62,327). The average ratings ranged from 2.7 to 4.5 out of 5 stars (mean 3.7). Of the 6 apps, 5 were free, but 4 offered a paid upgrade or paid in-app features. One app provided mobile access to a paid sleep web program.

Most were designed for commercial purposes, except for SleepHealth, which is restricted for research. The apps varied in the number of functions; all included a sleep diary and graphical feedback. All apps provided personalized advice based on sleep data, an accelerometer-derived sleep tracking (5/6), data relating to sleep stages (4/6), and sleep hygiene education (4/6).

Table 3. Preferences in sleep app features (abridged; N=167).

Features	n (%)	High insomnia (Insomnia Severity Index 17+)	Low insomnia	*P* value
Importance of...				
Sleep diary and tracking				.32
Little to no importance	19 (11.4)	9 (5)	10 (6)	
Important	148 (88.6)	88 (53)	60 (36)	
Tracking of diet, exercise, and other lifestyle factors				.11
Little to no importance	60 (35.9)	30 (18)	30 (18)	
Important	107 (64.1)	67 (40)	40 (24)	
Sharing or comparing sleep data with friends or other users				.45
Little to no importance	138 (82.6)	82 (19)	56 (34)	
Important	29 (17.4)	15 (9)	17 (8)	
Sharing sleep data with your doctor				<.001
Little to no importance	51 (30.5)	20 (12)	31 (19)	
Important	116 (69.5)	77 (46)	39 (23)	
Linking to a wearable device (eg, Fitbit, Jawbone, and ActiWatch)				.58
Little to no importance	84 (50.3)	47 (28)	37 (22)	
Important	83 (49.7)	50 (30)	33(20)	
Being usable offline (without an internet connection)				.14
Little to no importance	54 (32.3)	27 (16)	27 (16)	
Important	113 (67.7)	70 (42)	43 (36)	
Not requiring the phone to be by the bed				.49
Little to no importance	110 (65.9)	66 (40)	44 (26)	
Important	57 (34.1)	31 (19)	26 (16)	

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of selected apps.



Sample

For each app, the 100 most recent reviews were included except if there were fewer than 100 reviews (3 out of 6 apps), resulting in 434 reviews. Reviews were excluded if contextually irrelevant (eg, referring to weight loss features). A total of 385 reviews were sampled and 494 segments of text were coded (Multimedia Appendix 5 for full table).

Interrater Reliability

The interrater reliability between coders for the general comment theme was initially moderate (kappa=0.51). Following resolution of discrepancies, very high levels of interrater agreement were observed (kappa=0.99).

Content

Approximately half (66/122, 54%) of *content* mentions were praise for the app. App features were most frequently commented on (100/122, 82%). The alarm was the most salient *app feature* (43/122, 35%), attracting slightly less praise (15/43, 35%) than critique (28/43, 65%). The sleep diary was the next most salient *app feature* with 9% (11/122) of mentions, most of which were praise (6/11, 55%). Informative content was the most commonly mentioned *information* aspect (15/122, 12%) of which most mentions were praise (12/15, 80%; Figure 3).

Functionality

Data synchronization was the most commonly mentioned functionality (40/135, 30%), the majority of which were critiques (29/40, 73%). Bugs were the second most frequently mentioned (41/135, 30%). Mentions of data accuracy followed (29/135, 21%), which included a high proportion of critiques (24/29, 83%; Figure 4).

Figure 3. Bar graph presenting frequency of mentions for app content.

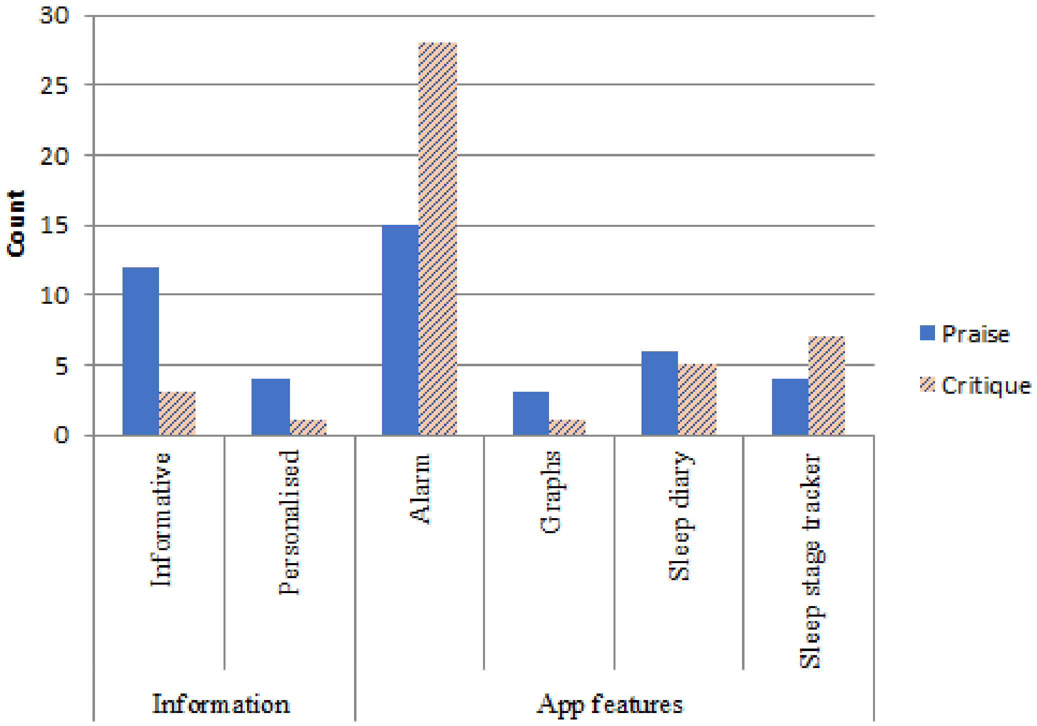
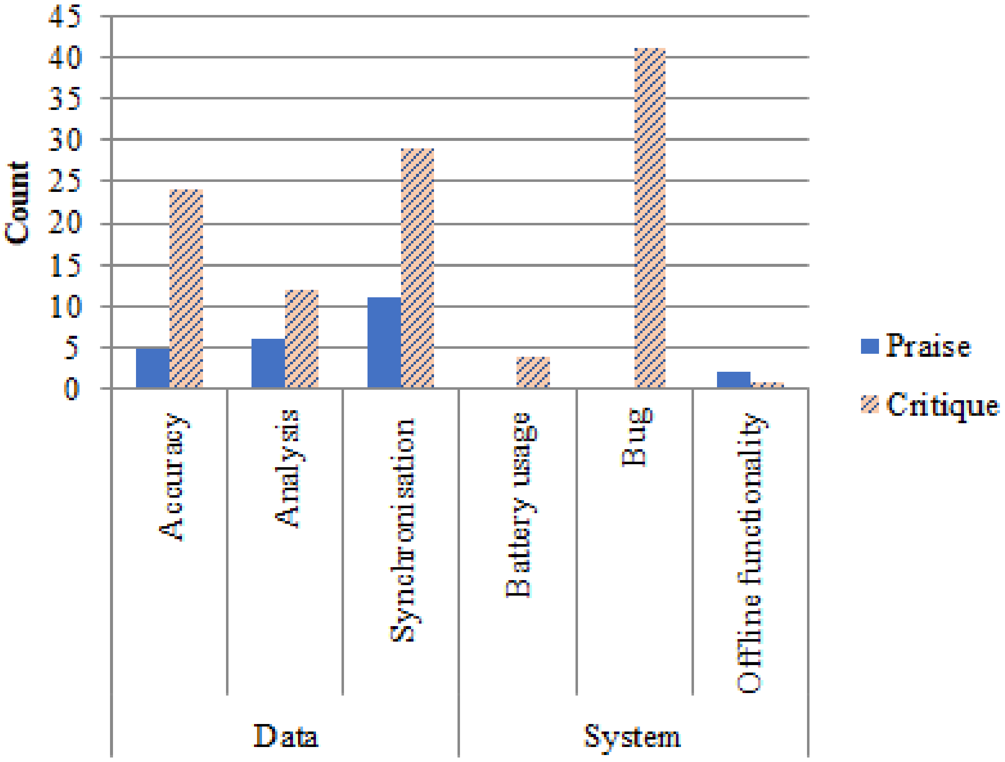


Figure 4. Bar graph presenting frequency of mentions for functionality.



User Experience

Apps' effectiveness received the highest proportion of *user experience* mentions (29/51, 57%). The general design or esthetics received the second highest proportion of mentions (13/51, 25%).

Other

App cost received over half of the *other* mentions (32/53, 60%), which were mostly critiques. In-app purchases were the second most frequently critiqued (11/53, 21%).

Discussion

Overview

This study explored the user preferences and needs of individuals with poor sleep or insomnia symptoms to guide the future development of a sleep mobile app. We found most participants expressed interest and willingness to use mobile apps for sleep. The triangulation of results was largely consistent, revealing that sleep app content and functionality are the most important components affecting user experience, which is consistent with previous findings [20,22-24].

Common Themes/Universal to mHealth Apps

Several generalizable themes emerged while exploring mHealth apps relating to sleep problems. This was in addition to themes common across app development where engagement is maintained by usability, intuitiveness, and bug-free designs [23-26]. Bugs were the most commonly reported problem. App reviewers criticized bugs affecting the app content and functionality. They emphasized the potential utility of these features that was impeded by the bug.

In both our focus groups and app reviews, the value of good quality information related to sleep health was most important. Previous studies have similarly highlighted the significance of relevant informative content that is central and accessible within digital treatments for cystic fibrosis [27] and harmful drinking in young adults [20]. We found that within our focus group sample, the type and extent of information in digital treatments can have counterproductive emotional and motivational consequences. Feelings of condescension from tedious information have also been found in cohorts with eating disorders [28,29] and healthy young adults [30], where content and reminder features are perceived as didactic. These responses may prompt individuals to question continuation and engagement with an app. In light of this, app developers need to ensure educational content is presented in a relevant and meaningful manner and tested with actual end users.

Interestingly, these feelings of belittlement were not echoed by the app reviews. This is perhaps due to contrasting sample characteristics between app reviews from the general public and focus group participants. Our long-term insomnia sufferers in the focus groups were recruited from a clinic and had prior exposure to basic sleep health information delivered in cognitive behavioral therapy for insomnia (CBTI) programs. A recent qualitative study revealed that many insomnia patients believed they knew about and had already attempted sleep hygiene practices [31]. They appeared to switch off when being readvised these from health professionals [31]. Further, given that our focus groups were exploring both app and web resources, the generally repetitive content of Web-based CBTI programs which is used to reinforce learnings may also be perceived as demeaning.

Some focus group participants found existing content to be complex and overwhelming. Resources expected impractical time commitments without a clear and pragmatic rationale for their participation. A delicate balance is necessary between content that is not too basic and perceived as condescending and being too complicated and overwhelming. Dennison et al highlighted how the challenge of burdensome and effortful apps can cause users to disengage [30]. Instead of text-heavy information, the answer may lie in visual and interactive content, conversational interfaces, and offering the user more autonomy, for example, to skip or drill down to more detailed information by choice. This sense of autonomy in users has shown to be important for motivation and engagement [32].

These techniques allow users to personalize content according to their needs which participants highlighted as an important requirement. Personalization also meant different things to different users. Our focus group participants stressed the importance of information personalization; the program adapting to an individual's symptoms while survey open-ended questions suggested customizability of app features, for example, relaxing sounds and meditation voices were desirable. Given the considerable heterogeneity in preferences, an innovative approach to personalization may involve an initial questionnaire to assess certain design requirements as well as users' severity and chronicity of sleep disturbance and filter app content accordingly. This customization ensures users are receiving meaningful information that can help sustain engagement without either trivializing or over complicating the user journey.

Themes Specific to Sleep Apps

Across methods, there were commonly desired app features specific to sleep, for example, alarm clocks. Given that most individuals already utilize the alarm clock in their mobile phones (68%), leveraging this in the design of an app may promote user engagement. Although the alarm is a salient app feature, it attracted more critique than praise in the app reviews, usually due to bugs, suggesting it may detract from a user's experience and retention when poorly implemented. The focus group participants added that an ideal alarm should go beyond the basic functionality of waking from sleep, involving customizable functionality with creative and sophisticated methods to wake you up and ensure you stay awake. Interestingly, this is in conflict with clinical practice which advises against distractions (including mobile phones) during bedtime [33]. Where major differences exist between user and clinical requirements, further research is required in reaching a compromise between evidence-based aspects and app features that engage.

The second major theme specific to sleep mobile applications was the ability to track sleep quantity and quality. Both focus group and app review participants were interested in synchronization with a WD to track and review their sleep. In addition to the advantages of WD tracking such as objective sleep metrics and sleep stage visualization, WD synchronization may increase usability by reducing burden related to manual sleep data entry. Interestingly, the importance of sleep tracking was echoed by only half of the survey sample. This discrepancy may reflect a selection bias inherent in the app review sample as all selected apps included WD synchronization, compared with only one-third of the survey sample being WD users. Correspondingly, a secondary analysis (not reported in the results) showed that WD users from the survey sample were more likely to indicate the importance of WD synchronization. Of these WD users, sleep tracking was identified as one of the most commonly used WD feature, consistent with previous research [34]. Given the anticipated proliferation in the global health care WD market [35], the addition of WD linkage is vital to the development of sleep-related mobile apps.

In addition to sleep tracking, participants suggested that subjective input supplementary to objective data can help data interpretation. Specifically, both focus group and survey users were interested in recording subjective lifestyle factors tracked against their sleep. In line with our findings, previous studies have shown similar needs for tracking contextual factors that assist in data interpretation to identify potential causes for poor sleep [31,36]. The integration of both subjective and objective sleep data tracking is important in supporting data interpretation and thereby enhancing engagement.

Another feature that may be offered with customization options is the ability to share sleep data with clinicians. Our results showed that those with high insomnia severity (ISI ≥17) were more likely to consider this function valuable in a sleep app. This is consistent with research showing the propensity for health care utilization is greater in those suffering from increased severity of sleep disturbance [37]. However, it is unclear whether clinicians are ready to use WD sleep data in clinical practice owing to a paucity of validation with traditional sleep technologies (eg, polysomnography). Given its potential utility, the guidance of sleep clinicians would be beneficial in determining which sleep data are of clinical value.

Limitations

Our study has several limitations. First, those who agreed to participate in the focus group and survey were potentially more technology literate and so may not help the development of apps seeking to attract new users. Similarly, app reviewers are likely to be confident technology users given their engagement in a public forum. Despite this probable response bias, the survey sample consisted of a wide age range and varying levels of experience with health apps, providing a good representation of current mobile phone owners who would be potential app users.

Given there was a gender imbalance between the survey sample (88% female) and a male-dominated focus group sample, and we do not know the gender mix of the app reviewers, this may explain some difference between these method’s outcomes. This may reduce the generalizability of our findings.

In addition, we cannot ascertain the extent of sleep disturbance in the app reviewers. Although all apps aimed to improve sleep and only sleep-relevant reviews were analyzed, one app addressed health more broadly and potentially attracted a diverse sample including healthy sleepers.

It is also worth noting that app reviews for previous app versions in the App Store are automatically removed once developers release a new version. Therefore, the study represents a snapshot of app reviews for sleep apps as of February 2018.

Conclusions

mHealth apps are usually posited at early intervention as either standalone or part of stepped-care programs; however, both require apps with sufficient engagement to enable therapeutic effects. To inform the development of a sleep mobile app, a user-centered approach is adopted to understand the needs and preferences that drive individuals with sleep disturbances to engage with sleep apps. Our findings highlight the importance of app content and functionality. In particular, developers should ensure the personalization of app content, including the customization of app features and information. Informative content should be targeted to the users’ individual sleep experience (the nature and extent of disturbance). Furthermore, developers need to consider including a sleep diary and alarm as a minimum but bear in mind that users seek more sophistication with customizable functionality and sleep diaries with subjective and objective WD tracking. Although further research is required to determine the efficacy of sleep apps, employing user-centric designs to develop engaging apps that are evidence based offers a new opportunity to advance clinical care for sleep.

Acknowledgments

This research is supported by the CRC for Alertness, Safety, and Productivity.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Online survey.

[PDF File (Adobe PDF File), 245KB - mental_v6i5e13895_app1.pdf]

Multimedia Appendix 2

Full coding scheme.

[PDF File (Adobe PDF File), 135KB - mental_v6i5e13895_app2.pdf]

Multimedia Appendix 3

Main themes and subthemes from workshops.

[PDF File (Adobe PDF File), 114KB - mental_v6i5e13895_app3.pdf]

Multimedia Appendix 4

Characteristics and features of sleep apps in the sample.

[PDF File (Adobe PDF File), 132KB - mental_v6i5e13895_app4.pdf]

Multimedia Appendix 5

Full table of coding frequencies by category.

[PDF File (Adobe PDF File), 113KB - mental_v6i5e13895_app5.pdf]

References

1. Adams RJ, Appleton SL, Taylor AW, Gill TK, Lang C, McEvoy RD, et al. Sleep health of Australian adults in 2016: results of the 2016 Sleep Health Foundation national survey. Sleep Health 2017 Feb;3(1):35-42. [doi: 10.1016/j.sleh.2016.11.005] [Medline: 28346149]
2. Roth T. New developments for treating sleep disorders. J Clin Psychiatry 2001;62(Suppl 10):3-4 [FREE Full text] [Medline: 11388588]
3. Rosen RC, Zozula R, Jahn EG, Carson JL. Low rates of recognition of sleep disorders in primary care: comparison of a community-based versus clinical academic setting. Sleep Med 2001 Jan;2(1):47-55. [Medline: 11152982]
4. Research2Guidance. 2018. 325,000 mobile health apps available in 2017 - Android now the leading mHealth platform URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2019-01-07] [WebCite Cache ID 75EwurraK]
5. Shin JC, Kim J, Grigsby-Toussaint D. Mobile phone interventions for sleep disorders and sleep quality: systematic review. JMIR Mhealth Uhealth 2017 Sep 7;5(9):e131 [FREE Full text] [doi: 10.2196/mhealth.7244] [Medline: 28882808]
6. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
7. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
8. Denzin NK. Triangulation: a case for methodological evaluation and combination. In: Sociological Methods. New York: McGraw-Hill; 1978:339-357.
9. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [Medline: 2748771]
10. Morin CM. Insomnia: Psychological Assessment and Management. New York, NY: Guilford Press; 1993.
11. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [Medline: 11438246]
12. Sanders E, Stappers P. Convivial Toolbox: Generative Research for the Front End of Design. Amsterdam: BIS Publishers; 2013.
13. Muller MJ. Participatory design: the third space in HCI. In: Jacko JA, Sears A, editors. Human-Computer Interaction: Development Process. Florida, United States: CRC Press; 2003.
14. Gray D, Brown S, Macanufo J. Gamestorming: A Playbook for Innovators, Rulebreakers, and Changemakers. Sebastopol: O'Reilly; 2010.
15. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
16. Chan C, West S, Glozier N. Commencing and persisting with a web-based cognitive behavioral intervention for insomnia: a qualitative study of treatment completers. J Med Internet Res 2017 Feb 10;19(2):e37 [FREE Full text] [doi: 10.2196/jmir.5639] [Medline: 28188124]
17. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]

18. Schilling J. On the pragmatics of qualitative assessment. Eur J Psychol Assess 2006 Jan;22(1):28-37. [doi: 10.1027/1015-5759.22.1.28]
19. Pagano D, Maalej W. User feedback in the appstore: an empirical study. 2013 Jun Presented at: IEEE International Conference on Requirements Engineering; July 15-19, 2013; Rio do Janeiro, Brazil p. 125-134.
20. Milward J, Khadjesari Z, Fincham-Campbell S, Deluca P, Watson R, Drummond C. User preferences for content, features, and style for an app to reduce harmful drinking in young adults: analysis of user feedback in app stores and focus group interviews. JMIR Mhealth Uhealth 2016 May 24;4(2):e47 [FREE Full text] [doi: 10.2196/mhealth.5242] [Medline: 27220371]
21. Thüring M, Mahlke S. Usability, aesthetics and emotions in human-technology interaction. Int J Psychol 2007 Aug;42(4):253-264. [doi: 10.1080/00207590701396674]
22. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
23. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 7;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
24. Khalid H, Shihab E, Nagappan M, Hassan AE. What do mobile app users complain about? IEEE Softw 2015 May;32(3):70-77. [doi: 10.1109/MS.2014.50]
25. Frie K, Hartmann-Boyce J, Jebb S, Albury C, Nourse R, Aveyard P. Insights from Google Play Store user reviews for the development of weight loss apps: mixed-method analysis. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e203 [FREE Full text] [doi: 10.2196/mhealth.8791] [Medline: 29273575]
26. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
27. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mHealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
28. Pollard CM, Howat PA, Pratt IS, Boushey CJ, Delp EJ, Kerr DA. Preferred tone of nutrition text messages for young adults: focus group testing. JMIR Mhealth Uhealth 2016 Jan 19;4(1):e1 [FREE Full text] [doi: 10.2196/mhealth.4764] [Medline: 26787115]
29. Lindgreen P, Lomborg K, Clausen L. Patient experiences using a self-monitoring app in eating disorder treatment: qualitative study. JMIR Mhealth Uhealth 2018 Jun 22;6(6):e10253 [FREE Full text] [doi: 10.2196/10253] [Medline: 29934285]
30. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
31. Middlemass J, Davy Z, Cavanagh K, Linehan C, Morgan K, Lawson S, et al. Integrating online communities and social networks with computerised treatment for insomnia: a qualitative study. Br J Gen Pract 2012 Dec;62(605):e840-e850 [FREE Full text] [doi: 10.3399/bjgp12X659321] [Medline: 23211265]
32. Peters D, Calvo R, Ryan R. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9(797) [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
33. Adams SK, Kisler TS. Sleep quality as a mediator between technology-related sleep quality, depression, and anxiety. Cyberpsychol Behav Soc Netw 2013 Jan;16(1):25-30. [doi: 10.1089/cyber.2012.0157] [Medline: 23320870]
34. Research2Guidance. User Survey: Wearables UK and US URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2019-01-07] [WebCite Cache ID 75EwurraK]
35. Grand View Research. 2018. Wearable Sensors Market Size, Share & Trends Analysis Report By Sensor Type, By Device (Smart Watch, Fitness Band, Smart Glasses, Smart Fabric), By Vertical, By Region, And Segment Forecast, 2018 - 2025 URL: https://www.grandviewresearch.com/industry-analysis/global-wearable-sensor-market? [accessed 2019-03-01] [WebCite Cache ID 76XNEa4BE]
36. Liu W, Ploderer B, Hoang T. In bed with technology: challenges and opportunities for sleep tracking. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction. Dec. 2015 Presented at: OzCHI'15; December 7-10, 2015; Melbourne, Australia p. 142-151. [doi: 10.1145/2838739.2838742]
37. Kapur V, Redline S, Nieto F, Young T, Newman A, Henderson J. The relationship between chronically disrupted sleep and healthcare use. Sleep 2002;25(3):289-296. [Medline: 12003159]

Abbreviations

CBTI: cognitive behavioral therapy for insomnia
ISI: Insomnia Severity Index
mHealth: mobile health
PSQI: Pittsburgh Sleep Quality Index

WD: wearable device

Edited by J Torous; submitted 04.03.19; peer-reviewed by V Rocío, A Powell; comments to author 19.03.19; revised version received 24.04.19; accepted 28.04.19; published 24.05.19.

Please cite as:
Aji M, Gordon C, Peters D, Bartlett D, Calvo RA, Naqshbandi K, Glozier N
Exploring User Needs and Preferences for Mobile Apps for Sleep Disturbance: Mixed Methods Study
JMIR Ment Health 2019;6(5):e13895
URL: http://mental.jmir.org/2019/5/e13895/
doi:10.2196/13895
PMID:31127714

Original Paper

How Is the Caregiver Doing? Capturing Caregivers' Experiences With a Reflective Toolkit

Lilian Bosch[1], MA; Marije Kanis[1], PhD; Julia Dunn[2], MSc; Kearsley A Stewart[2], PhD; Ben Kröse[1], PhD

[1]Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[2]Duke University, Durham, NC, United States

Corresponding Author:
Marije Kanis, PhD
Amsterdam University of Applied Sciences
Wibautstraat 2-4
Amsterdam, 1091 GM
Netherlands
Phone: 31 21156886
Email: m.kanis@hva.nl

Abstract

Background: This paper describes the *Co-Care-KIT*, a reflective toolkit designed to provide insights into the diverse experiences of home-based informal caregivers during the delivery of care to a relative or loved one.

Objective: The aim of this study was to evaluate the toolkit, including a custom-designed journal, tools for photography-based experience sampling, and heart rate tracking, which enables caregivers to collect and reflect on their positive and negative daily experiences *in situ*.

Methods: A 2-week field study with informal caregivers (N=7) was conducted to evaluate the Co-Care-KIT and to capture their daily personal emotional experiences. The collected data samples were analyzed and used for collaborative dialogue between the researcher and caregiver.

Results: The results suggest that the toolkit (1) increased caregivers' awareness of their own well-being through *in situ* reflection on their experiences; (2) empowered caregivers to share their identities and experiences as a caregiver within their social networks; (3) enabled the capturing of particularly positive experiences; and (4) provided caregivers reassurance with regards to their own mental health.

Conclusion: By enabling capturing and collaborative reflection, the kit helped to gain a new understanding of caregivers' day-to-day needs and emotional experiences.

(JMIR Ment Health 2019;6(5):e13688) doi:10.2196/13688

KEYWORDS

human-centered design; informal caregivers; experience sampling; home care; positive psychology; mental health; reflective technologies; well-being; HCI (human computer interaction)

Introduction

Longer life expectancies and health care reforms in professional health care provision have increased the reliance on informal caregivers (eg, family, friends, or neighbors), who carry out long-term in-home care of their relative or loved one [1,2]. Across the Organization for Economic Co-operation and Development countries, such as France, the United Kingdom, the United States, and the Netherlands, informal caregivers have become the backbone of long-term care provision. Over 1 in 10 adults provide informal (unpaid) care to their parent (36%), spouse (31%), friend (18%), or relative (18%) [2]. Providing care can impact caregivers' health and well-being in numerous ways. Understanding the diverse needs of informal caregivers is critical for finding ways for supporting their well-being. This is not only beneficial for them but also for the care recipients who often prefer someone they know well to provide (long-term) care for them [2]. Formal care providers should also have an interest in improving the well-being of informal caregivers, as informal caregivers play an imperative role in providing higher quality home-based care [3].

Technology has considerable potential to support home care provision. A large body of research exists in human-computer interaction (HCI) to mainly support patients [4-7] but also caregivers [6,8-13] in the context of in-home care. Thus far, however, in developing technologies for caregivers, research has generally emphasized functional support for caregiving (how can a caregiver provide better, more effective care?), whereas few studies focus on the well-being of caregivers [8,14,15]. Yet, researchers in HCI have already urged for moving and evolving from an efficiency approach toward a focus on user experience and toward finally embracing positive psychology approaches and focusing on wellness [16-20]. Similarly, in the development of technologies for home care, researchers in this field have already indicated the need for shifting focus from solutions for managing care to solutions for wellness [21,22]. The demand for and reliance on informal caregivers is likely to increase even more. Thus, it is now time to also acknowledge the remaining gap in technological approaches and solutions for positively identifying and supporting the caregiver experience and focus on how technological methods could place more importance on caregivers' well-being.

In 2013, the National Alliance for Caregiving [14] already stated in a timely report how technology can play a more meaningful role in helping informal caregivers:

> *There is an ongoing need for research on family caregivers, especially as technology dramatically impacts caregiving. More current, thorough and accurate data is needed about the diversity of caregiver roles and responsibilities, about what caregiving involves day-to-day and the nature of the burden it represents, and how much it impacts those around the caregiver.*

Better identifying and understanding the varying, complex, and personal experiences of caregivers in the contexts in which it takes place is crucial in exploring how (technological) solutions could improve the needs and well-being of caregivers. To address this, this study investigated how methods from HCI and design can help capture and understand the day-to-day (emotional) experiences of caregivers. For this purpose, a reflective study toolkit was designed. The Co-Care-KIT allows for caregivers to collect meaningful data about the personal, everyday experience of caregiving. The Co-Care-KIT serves 2 goals. First, to investigate how different tools for capturing caregivers' experience combined can work during a 2-week field study. Second, to gain in-depth knowledge about the experience of diverse caregivers through the qualitative data that were collected with the toolkit. The primary purpose was thus to evaluate a toolkit for better understanding the experiences of caregivers. This was done so to meet the ultimate ambition of better supporting caregivers' well-being. Consequently, by encouraging self-reflection, the use of the toolkit was designed to also be of value for the caregiver. A positive design approach was undertaken in trying to maximize the toolkit's use for the participants by facilitating caregivers to collect and reflect on experiences that were meaningful to them as these occurred in the wild.

The rest of this paper presents the design of the Co-Care-KIT toolkit based on studies of relevant methods from HCI and design for capturing experiences. It describes a 2-week field study with the Co-Care-KIT that was conducted with home-based informal caregivers in diverse care situations (N=7). Then it reports the study's findings, including an evaluation of the toolkit method. The paper concludes with a discussion of the lessons learned using this method. The main contributions of this paper constitute the design and evaluation of the rich toolkit; the lessons learned and new perspectives gained through deploying this reflective toolkit method; and findings concerning the emotional experience of caregivers from the rich data that were collected through this study. These findings build on and contribute to the research on how HCI can better understand and support informal caregiving.

Capturing Experiences

Informal Caregiving Research

In investigating the experience of caregivers, research has focused primarily on the more negative aspects. It has focused particularly on the burden that caregivers endure, such as psychological stress; depression; poor health; guilt about not doing enough; work and employment problems; and social effects such as isolation [23-28]. Nonetheless, caring for a loved one can also provide a sense of gratification to caregivers. The lack of attention paid to the positive aspects of caregiving has skewed perceptions of the experience of caregivers in research [29]. However, there are also positive aspects that have been highlighted in the literature [28-32]. Recently, scholars have started to emphasize on how positive aspects of caregiving can actually lower levels of reported depression of caregivers [31]. Conceptualizations of positive aspects of caregiving include the following: role satisfaction (a sense of satisfaction in performing tasks); emotional rewards (feeling appreciated or successful); personal growth (growing as a result of the role); competence and mastery (sense of mastery in the new role); relationship gains; sense of duty; and reciprocity (satisfaction in fulfilling a sense of duty) [31]. Nonetheless, every care situation is unique and different factors can contribute to the experience of caregivers. For example, the care load of an informal caregiver often depends on the constantly changing needs of their care recipient. It also depends on the presence (or absence) of other informal and formal caregivers. Moreover, whether the helping role of a caregiver causes them stress or a feeling of satisfaction can quickly change based on time and context [33]. Studies have predominantly used quantitative measures to investigate the experience of caregivers. These do not capture the contextual features that determine the experience of a caregiver [23]. As the experience of caregiving is not a fixed state but a dynamic process that depends on many factors, it is crucial to recognize these nuances to be able to support the well-being of caregivers. Thus, in investigating the experience of caregivers, dynamic factors such as time and context should be considered in an ongoing assessment of how a caregiver is doing from moment to moment and day to day. Moreover, caregivers may not be aware of or focused on their own health and well-being, as previous research has shown that caregivers prioritize the well-being of the person they care for over their own [8,34]. Thus, in studying the experience of caregivers, it is important to also encourage caregivers to reflect on their personal well-being and needs and provide tools to do so. In the next section, this paper explores how methods from HCI and design can help to capture and understand the day-to-day experiences of caregivers.

Methods

Experience Research in Human-Computer Interaction and Design

In HCI, there are a number of methods that involve capturing, measuring, and tracking daily habits and experiences. For example, Quantified Self technologies, such as sensing and lifelogging systems (the automatic capture of activities with digital tools such as wearables), have demonstrated potential in collecting bodily and day-to-day status data of people. Although there are some notable exceptions, such as the Three Good Things method that encourages people to record positive experiences [16,35], lifelogging and sensing activities often seem to focus on detecting the more negative emotional, functional, and bodily experiences, such as anxiety, stress, and depression [36-40]. However, these technologies can also be deployed for more positive purposes. For example, Sellen and Whittaker [41] have discussed how lifelogging could potentially support learning through later reminiscence (recollection of past events) and reflection (reflecting on and reviewing of past events). In psychology research, reflection is increasingly correlated with mental health and well-being [42,43]. Consequently, in HCI studies, there is also a growing interest in design for reflection [35,44-48], often involving participants to self-track or capture data. For example, Isaacs et al [35] have developed a lifelogging app that facilitates users to record and reflect on memories, ultimately aiming to increase the well-being of its users. Moreover, diary studies [49] are now more widely used in HCI to evoke reflections on participants' views and experiences *in situ*. In addition, the expression of emotions through expressive writing is a common therapy for improving a person's well-being [50-52]. Furthermore, cultural probes [53] are materials used in design research to elicit expression and responses from users. For example, in a notable study in 2016 with informal caregivers, probes (in the form of diaries and photo cameras) were used as tools to gather stimulus for interviews [54]. Probes can come in many forms, such as diaries, postcards, cameras, objects, or games, and are typically combined in toolkits to enable rich and varied qualitative inquiry from prospective users [55]. Diary methods have been used in research to investigate social, psychological, and physiological processes within daily situations, and avoid some of the recall biases inherent to single report interviews [56]. Interpretation of data gathered through design probes is often mostly done by designers and researchers to ultimately inspire the (co-)design process of new products or services. Thus, using probes for collaborative interpretation and self-reflection by the target group has largely been uncharted territory so far. Another method that is specifically developed for measuring the experience of people is the *Experience Sampling Method* (ESM). ESM [57] originates from the field of positive psychology that places focus on the positive aspects of life and where it is commonly used for assessing *in situ* subjective happiness of people by having them report on their thoughts and feelings at signaled moments throughout the day [58]. It combines the ecological validity of naturalistic behavioral observation with elements of a diary and the precision of questionnaire measures [57]. This approach has also further been developed for the use with mobile devices and various media, such as photography, to capture data and experiences from users [59-61]. Photo elicitation allows for the development of rich phenomenological data to research the lived experience and provides subjects an opportunity to express their experience from their own perspective [62]. The process has been described as highly empowering because it offers an opportunity for research participants to actively engage in the research process [63]. ESM has an advantage over other research methods that also collect information about the context and lives of people (such as observations, interviews, or diary studies), namely, it enables direct capturing of experiences in the moment they occur. The Atlas of Caregiving [14] also stresses the need for experiential and reflective approaches toward the daily practices of caregiving but focuses on a different cultural (the United States) and social context (family).

Design for Reflection

Baumer [64] coined the term *reflective informatics* and distinguishes between 3 dimensions (perspectives) to inform and inspire how design can potentially support reflecting: first, there is *breakdown* (eg, when a system suddenly does not perform as a user expects it to, an opportunity for reflection is created); second is *inquiry* (supporting reviewing of past experiences such as with reflecting on past journal entries); and the final dimension is *transformation* (not just examining the current state of something but envisioning alternatives and supporting change).

As this paper focuses on capturing and understanding the personal and varying daily experience of caregivers (or inquiring into the experience of caregiving), this study fits mostly into Baumer's dimension of *inquiry*. We address this by enabling caregivers to self-capture experiences that are meaningful to them and, in doing so, facilitate and empower caregivers to reflect on their current and past experiences.

This study investigated methods from HCI, (reflective) design, and (positive) psychology for capturing the experience of people. Such methods have not yet been studied vastly with caregivers. Moreover, these methods focus on capturing samples by participants but often leave the interpretation of data to the researchers. Therefore, these investigations formed the basis for the design of a reflective study toolkit for caregivers: the Co-Care-KIT. This study kit contains a variety of tools that combined enable caregivers to capture and reflect on their varying, contextual (emotional), and personal experiences as these occur *in situ*. The tools not only enable capturing of (positive) experiences but make it possible to visualize these. In this way, the caregivers' experiences can be discussed in a dialogue between the researchers and caregivers. The next section describes the design and development process of the Co-Care-KIT.

Co-Care-Kit Design

The idea of a toolkit for caregivers was also inspired by a study by Bosch and Kanis [8], which identified several design opportunities for caregivers. This study also highlighted how capturing and cherishing moments during caregiving could emphasize the caregivers' positive outlook. Therefore, in the design of the individual tools in the Co-Care-KIT, emphasis was placed on eliciting reflections about how the caregiver was doing both in the moment and generally. A variety of complementary methods of elicitation were included to understand the caregiver experience fully. The final Co-Care-KIT toolkit (see Figure 1) comprises the following tools:

1. *A custom-designed journal* for writing and reflecting about daily experiences. In addition to recording daily thoughts and experiences, the journal contained predesigned pages for supporting experience disclosure.
2. *An iPhone with an app for photography-based experience sampling:* this app enabled caregivers to visually capture their experiences and rate how they are feeling at signaled moments throughout the day.
3. *A heart rate tracker wristband* that caregivers can use to continuously measure their heart rate variability (HRV), as an indication of their stress.
4. *Supplementary material:* Online and offline informative resources, such as bookmarks and leaflets, on caregiving support were included in the kit. Finally, additional writing aids to use with the notebook were provided, such as pens, markers, and Post-it notes.

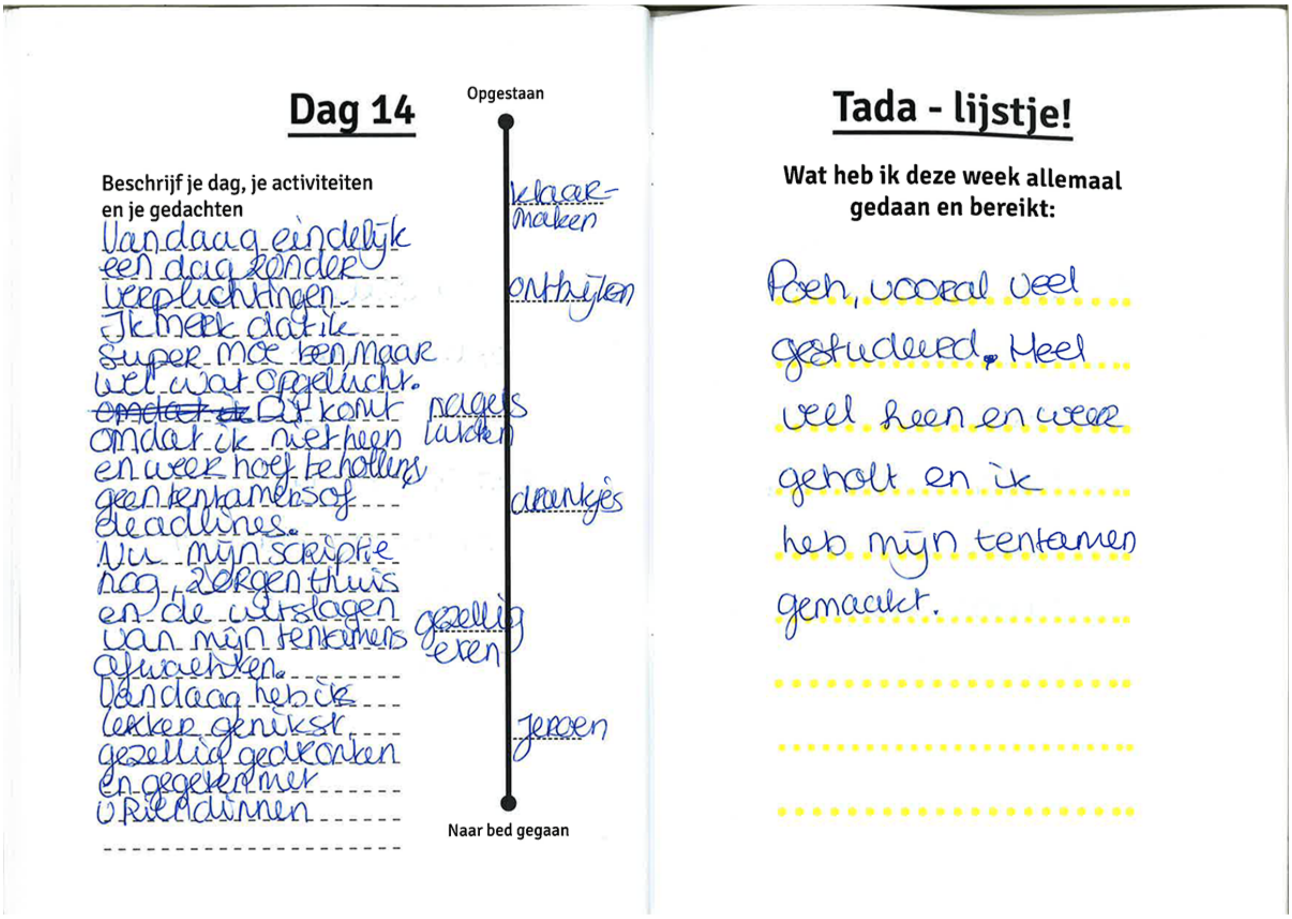
A basic and unpolished look was chosen for the visual design of the Co-Care-KIT, using materials such as cardboard and paper, to maximize its appeal to a wide variety of participants (both young and adult caregivers, male and female) and to attract participants in using the provided tools and materials. These tools were selected because these could complement each other, namely, appealing to different types of users and ways of (positively) identifying the mental and experiential state. A key goal was that the caregivers could directly benefit from participation. The kit, therefore, not only provided additional supplements but also provided tools that enabled self-reflection so as to support more awareness of their own needs and well-being.

Journal: Writing and Reflecting

The first tool in the Co-Care-KIT is a custom-designed journal in which the participants can report their activities in their own time during the study. To support this process, the journal includes predesigned pages (see Figures 1-3), for example, some contain a question or a beginning of a mind map about topics such as "positive emotions I am feeling. . .", "what new skills have you learned from the care you provide?", or "what do you do to relax?" to provoke reflection about diverse (emotional) experiences; other pages include daily timelines on which participants can note their daily activities.

Figure 1. The Co-Care-KIT, including a custom-designed journal, photography-based experience sampling app, and heart rate tracker.

Figure 2. Journal pages from P5. On the left page: a diary of day 14 of the field study and on the right page: a list of things she was proud to have accomplished this week.



Dag 14

Beschrijf je dag, je activiteiten
en je gedachten

Vandaag eindelijk
een dag zonder
verplichtingen.
Ik merk dat ik
super moe ben maar
wel wat opgelucht.
~~omdat ik~~ Dit komt
omdat ik niet heen
en weer hoef te hollen,
geen tentamens of
deadlines.
Nu mijn scriptie
nog, zorgen thuis
en de uitslagen
van mijn tentamens
afwachten.
Vandaag heb ik
lekker genikst,
gezellig gedronken
en gegeten met
vriendinnen

Opgestaan
klaar-maken
ontbijten
nagels lakken
drankjes
gezellig eten
Jeroen
Naar bed gegaan

Tada - lijstje!

Wat heb ik deze week allemaal gedaan en bereikt:

Poeh, vooral veel
gestudeerd. Heel
veel heen en weer
geholt en ik
heb mijn tentamen
gemaakt.

Experience Sampling: Capturing Everyday Moments

Second, the Co-Care-KIT contains an iPhone equipped with a photography-based experience sampling app (see Figure 4). The app is programmed to signal participants at random intervals during the day to reflect on what they are doing and how they are feeling. This involves taking a photo and completing a short survey, including writing a caption of the captured moment and indicating how they feel on a 5-point happiness scale. Photo elicitation evokes emotions and feelings and forms a rich source of data. The alarms go off up to a maximum count of 6 times a day so as to capture the experience at different times and showcase a diversity of moments from caregivers over the course of a day. Experience sampling research indicates that a maximum of 6 alarms is considered appropriate to not overburden the participant [57].

Alarms can be set within a time frame in consultation with participating caregivers. Individual alarms can be either snoozed or ignored when these happen to go off at an inconvenient time.

Tracking How the Caregiver is Doing

Third, a heart rate monitor wristband (the Mio Link [65]) is included in the Co-Care-KIT to complement the journal entries and photography-based experience sampling data of participants with objective data about how they are doing during the study. This enables caregivers to track their heart rate using an HRV tracking app [66] on the iPhone. HRV is used to describe a number of measurements based on consecutive heartbeats and, in research, HRV measurements are used as a physiological indicator for recognizing mental effort and emotions, such as stress and worry [67-70]. The large body of research done by the MindGames research group [71] showed that biofeedback such as HRV can be creatively used for obtaining a positive state of mind. Current neurobiological evidence suggests that HRV is impacted by stress and supports its use for the objective assessment of psychological health and stress [72]. Nonetheless, measuring physiological data such as HRV with a noninvasive consumer wearable, such as the optical heart rate tracker wristband that was used, is not very reliable. For example, missing a heartbeat because of an incorrectly placed wristband could result in false measurements. However, accuracy was not the main goal but rather to have some type of objective, a physiological indicator to reflect the well-being of caregivers to evaluate if and how caregivers value such data. Therefore, this was nevertheless included in the Co-Care-KIT (with instructions). For this reason, caregivers were able to use the heart rate monitor during the study to track but were not able to see a visual representation of these data until after the field study.

Figure 3. Journal pages from P7 (top) and P4 (bottom). The statement at the top of the page prompts participants: People most involved in the care process. The participant is in the center of the page and writes down who is involved (care recipient, family, friends, and the professional help).

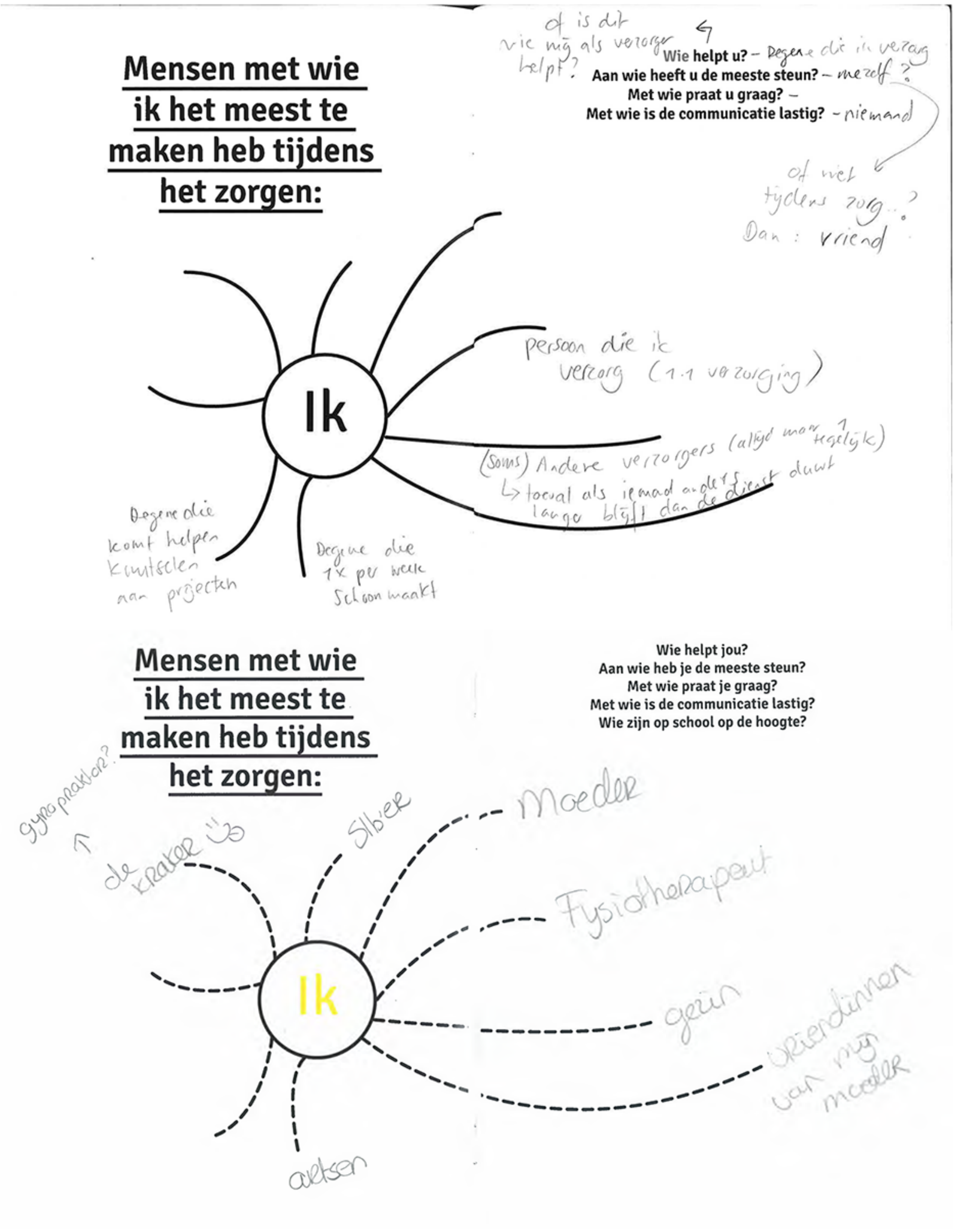


Figure 4. The experience sampling app: taking a picture, writing a caption, and indicating mood.

Neemt u a.u.b. een foto

Omschrijf hieronder kort de genomen foto:

Een uitleg...

Hoe gaat het met jou?

Providing Information

Online and offline informative resources for caregivers were included in the kit, as a previous study [8] indicated a clear need for this. These informative materials were obtained and selected through consulting Dutch informal care organizations and included leaflets about support groups for informal caregivers and information about informal caregiver organizations. Moreover, Web page bookmarks with information about informal care support services were preprogrammed on the iPhone.

Making Data Discussable

To be able to discuss the collected data samples with caregivers, we see it as critical to visualize each form of experience sample. Therefore, the heart rate measurements of each caregiver were plotted on daily timelines, including the photos that had been taken. This was done to create an overview and to show context data such as time, location, and people. Similarly, the pages in the journal were designed in such a way that these can be interpreted at a glance (see Figures 2 and 3). In summary, the Co-Care-KIT combined various methods for (subjective and objective) experience disclosure into 1 toolkit to enable caregivers to capture and *self-reflect* on their emotional experiences *in situ*. The collected experience samples then served as a rich source for dialogue between participants and researchers.

Study Approach

A 2-week field study with the Co-Care-KIT was conducted with home-based informal caregivers in diverse care situations (N=7). This field study was preceded by semistructured introductory interviews with each of the participants to gain an understanding about their general experience of caregiving, and each individual field study concluded with a closing interview (lasting approximately 1.5 hour) during which the data that were collected with the Co-Care-KIT were discussed and evaluated together with the participants.

Recruiting Informal Caregivers

The majority of studies with informal caregivers in HCI focus on caregivers of a specific patient group (eg, caregivers of people that are diagnosed with Alzheimer disease). Moreover, even though people of all ages serve as informal caregivers, the majority of studies are focused on adult caregivers [12,54]. This study goes beyond these specific caregiver or patient groups and focuses on capturing the caregiving experiences of diverse caregivers in diverse care situations. An intentionally wide variety of caregivers was approached for recruitment. The challenge in recruiting potential participants is that caregivers often fail to identify with the term *informal caregiver* because of the gradual assumption of the role [73,74]. Therefore, many different approaches were undertaken for recruiting study participants: collaborating with informal care organizations and research groups in the Netherlands and beyond; volunteering as a caregiver to connect with the target population; promotional flyers at a doctors' office, placing online calls in Facebook caregiver support groups; and reaching out to young caregivers who had already taken part in a university-wide study and who indicated their role and consented to be contacted for future research studies.

Study Design

After recruitment and obtaining formal and detailed consent, *semistructured introductory interviews* were conducted with informal caregivers of care recipients with wide-ranging care needs (N=7; see Table 1).

Table 1. Overview of study population and collected data.

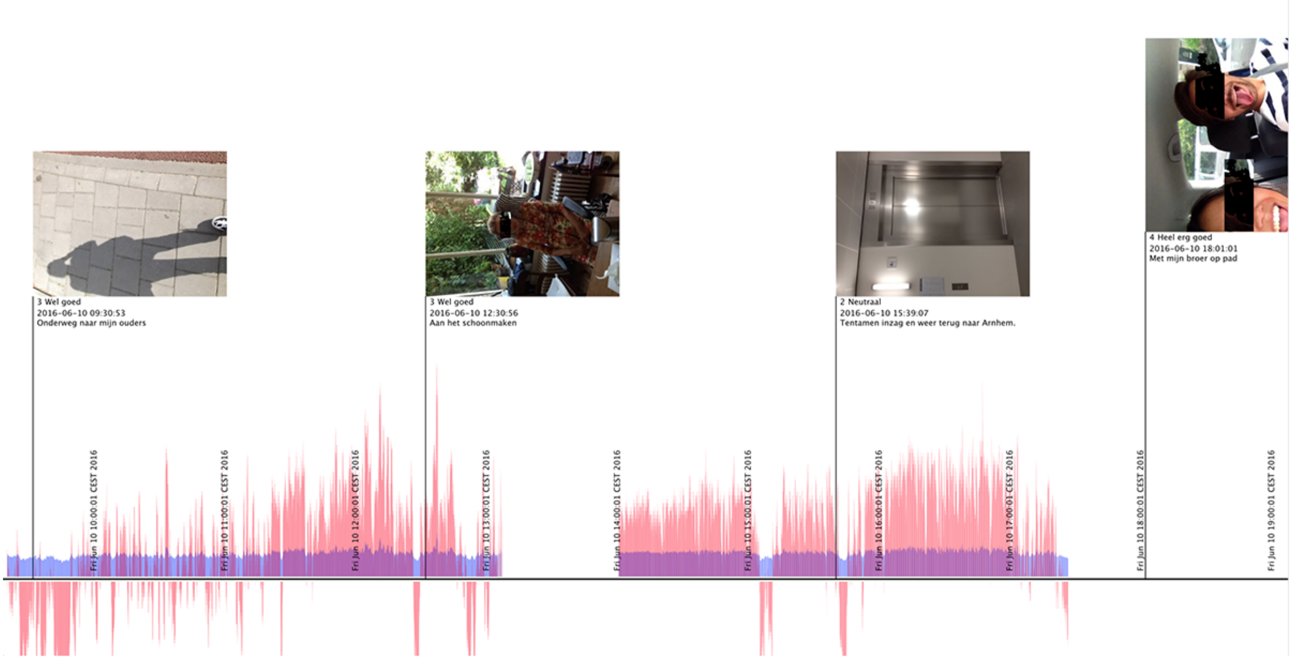
Participant	Gender	Age	Health concerns of care recipient	Working part-time (PT), full-time (FT), or studying	Relation to care recipient	Hours spent caregiving per week	Duration of field study (days)	Photos collected	Heart rate tracking (days)	Journal kept
P1	M	63	Dementia, osteoporosis, and visual impairment	PT	Son/son-in-law	3-8	8	30	7	Yes
P2	F	36	Stroke, paralysis	PT	Daughter	8-16	14	32	1	Yes
P3	F	35	Spinal cord injury	PT	Neighbor	<3	14	30	11	Yes
P4	F	57	Dementia, osteoporosis	PT	Daughter	8-16	11	42	3	Yes
P5	F	23	Amyotrophic lateral sclerosis	FT student	Daughter	3-8	12	17	1	Yes
P6	F	22	Old age, frailty	FT Student and PT working	Granddaughter	8-16	7	25	3	Yes
P7	F	26	Chronic pain		Daughter	8-16	14	23	5	Yes

Each caregiver was questioned about (1) *the context of their care situations* (eg, for how long they had been taking care of someone; what their relationship with this person was; why they cared for this person; and what their caregiving tasks entailed) and (2) *their (emotional) experiences of caring for someone* (eg, what they enjoyed most about caregiving and what they did not like about caregiving). An additional theme was discussed with young caregivers who were also enrolled in college, namely (3) *their experience of combining school and caregiving* (eg, how they combined school work, caregiving, and their personal lives; whether their school performance was suffering because of their caregiving work; and how the university could better support them). To gain an understanding of each caregiver's perceived burden, the 13-question Caregiver Strain Index (CSI) [75] was used during each interview. The CSI is computed by summing the 0 (no) and 1 (yes) responses for the 13 items. Therefore, CSI scores range from 0 to 13. Positive responses to 7 or more items on the index indicate a greater level of strain, and the instrument can be used with individuals of any age. It should be noted that the CSI instrument was originally developed to screen partners of older adults for caregiver strain. For the purpose of this study, a slightly modified version of the CSI was used with all participating caregivers: in one of the questions, the phrase *my partner* was changed to *my care recipient*. Interviews took place in the caregivers' home or at the university, and each conversation took approximately 1.5 hours including instructions. Owing to the small sample size, the CSI only provides a preliminary insight of participants in the study. Next, a *field study with the Co-Care-KIT study toolkit* followed with all of the participants. As informal caregivers form a study population that can already be quite burdened by their circumstances, the duration of the field study was determined in consultation with each participant and varied between a minimum of 7 days and a maximum of 2 weeks. The informal caregivers were provided with the Co-Care-KIT toolkit and each tool was explained and demonstrated. During the study, the researchers contacted each participant at least once to check if everything was going well and to keep the study motivation going. After the field study was completed, the electronic devices and collected experience data were returned to the researchers and a *final closing interview* was scheduled during which participants discussed and interpreted their captured experience data together with the researchers. In preparation for this interview, participants' heart rate data were visualized on daily timelines, which also showed the photos that they had taken (Figure 5). These visualizations served as stimulus for the interviews so caregivers could interpret and reflect on their experiences and what they had learned.

Data Analysis

All interviews (7 introductory and 7 final interviews) were transcribed and then coded by 2 researchers. The transcriptions were analyzed using applied thematic analysis, a broad inductive method in which key themes are identified in text and transformed into codes [76]. Applied thematic analysis as a method assumes a researcher's effect on the data. Before the final interviews, the journals that participants had kept during the field study were read and used to formulate additional questions. During the final interviews, the data that were collected with the study kit were thoroughly discussed with each caregiver by using the experience data samples as conversation props: the photos and heart rate measurements were spread out on the table and were discussed and interpreted by the caregivers while the researchers questioned them about it. This was the main method used for analyzing the qualitative data from the kit.

Figure 5. Heart rate measurements of P5 plotted on a 1-day timeline, including the affective state and pictures that were taken that day.



Results

A group of informal caregivers (N=7, 1 male and 6 females) were recruited for this study (see Table 1). The study population consisted of caregivers aged between 22 and 63 years, amid 3 young caregivers who were still in college.

Introductory Interviews

The study population included caregivers in diverse care contexts (see Table 1). Most participants cared primarily for their parents (n=5) and in other cases for their grandparents (n=1) or their neighbor (n=1) and cared for persons with diverse and sometimes multiple health issues, varying from Alzheimer disease, old age, and physical disabilities to chronic pain. Overall, caregivers acknowledged a positive outlook on their care situations. Participants seemed very open in discussing their personal experiences during the interviews and the majority of the conversations were perceived optimistic and positive, even while discussing difficult subjects and situations. As P2 explained: "I liked the study method, so that makes it easier to talk about the subject." However, despite the majority of participants liking the kit and having a seemingly positive outlook and experience of caregiving, their average measured burden was relatively high, namely, 6.7 (SD 2.98) on the CSI. This 13-question instrument is the most used scale for measuring subjective burden among caregivers. The CSI is computed by summing the 0 (no) and 1 (yes) responses for the 13 items. Therefore, CSI scores range from 0 to 13. Positive responses to 7 or more items on the index indicate a greater level of strain. For the adult participating caregivers, the measured CSI was a little lower and more varied, with an average of 6.0 (SD 3.9), than for young caregiver participants, whose average was 7.7 (SD 1.15). Caregivers found it difficult to indicate how much time they actually spent on caregiving on average, as care activities happened routinely throughout the day. Moreover, even though the caregivers had been recruited for a study about the experience of caregivers, they were seemingly trained in thinking about what is best for the person they care for: on multiple occasions during interviews, the researchers had to remind caregivers to answer questions from their own perspectives, instead of from their care recipients' perspective.

Field Study With the Co-Care-KIT

The participants used the Co-Care-KIT for 1 to 2 weeks each, with an average of 11 days per participant (see Table 1). Overall, the experience of using the toolkit was found positive and resulted in a considerable amount of experience data recorded by caregivers. In total, 199 photos were captured (mean 28 per participant), 31 days of heart rate measurements were recorded (mean 4.4 days per participant), and every participant used the Co-Care-KIT's journal during the study. The participants interpreted their captured data together with the researchers during the closing interviews. The findings reported next are based on these interpretations.

Photography-Based Experience Sampling

Many photos were collected, even though some participants reported that they sometimes skipped recording a sample during the study when consecutive alarms would go off and they were still in the same place or engaging in the same activity that had been recorded before. Signaling participants to capture photos at random moments throughout the day and initiating a small moment of reflection by writing a caption ensured that participants not only captured the highlights of the day but also the *mundane* experiences. This resulted in a varied and detailed visual report of their lives. As P1 stated,

> *We already take a lot of photos now, but those are the typical "oh look how nice" or "look how special" moments.*

Thus, experiences captured with the photo elicitation app were diverse: moments spent in hospital; performing care tasks; during cooking, cleaning, or doing laundry; time spent with family members or friends; at work; at school; during homework; or while traveling. The photos taken during the study revealed a lot about the daily context of each caregiver's life: their home and the places they visit, their family, the person they care for, and all the different things the caregivers do in a day on top of their care responsibilities.

During the closing interviews, each participant was asked to organize and discuss the photos they had captured in whichever way they preferred. The photos that participants most often selected as positive, unsurprisingly, involved activities during leisure time, spending time with friends or family, and cooking. Caregivers also indicated moments spent with the person they cared for as most special to them. Negative moments included times that the caregivers experienced worry or stress: such as during a hospital visit; when the care recipient was not doing well; or, for young caregivers, moments when they were stressing over their homework. Moreover, seeing repetitive photos was confronting for some participants because it emphasized how much time they spent inside the house (P4) or at work (P3). Some participants did not like performing care tasks, whereas others enjoyed them because it made them feel appreciated and helpful. For example, P6 (see Figure 6) stated:

> *My grandmother is usually really happy when I help her with her groceryshopping. She knows everyone, including me, has a lot on our plates, so when I take my time with her and don't rush her through the store she is really appreciative and that makes me feel good about it.*

Another participant (P5) described how some days were stressful when she had a lot of studying to do, but also had to help her parents with chores (Figure 6). However, there were also photos of moments that were not simply positive or negative and talking about these moments revealed some conflicting feelings a caregiver can experience that were not captured in the photo caption, emotion scale, or journal. For example, 1 caregiver (P6) photographed a sports game that she went to in the weekend and although she had categorized this moment under *relaxing activities*, she explained that during the moment she had felt guilty over not being able to visit her grandfather that day, which made it difficult for her to enjoy the moment. This example also demonstrates how caregivers, even when they are not actively carrying out care activities can still feel emotionally burdened.

Despite the mixed emotions, the photo sampling seemed to help with positive reflection, as P5 stated,

> *Most valuable was that I had to write a short comment about the photo. I liked that. From the photos I learned that I like where I am and I like what I do, during caregiving.*

Journal Keeping

All of the participants wrote in the journal throughout the study, although with varying detail. They found it was less immediate for capturing experiences than the photo elicitation app. For example, P1 stated,

> *The journal is less "in the moment". I wrote in it afterwards, based on memory, not during the day.*

Of them, 1 participant (P5) explained it was sometimes hard to be honest in her journal entries on days where she did not accomplish the things that she had planned because she was too tired. Nonetheless, she kept a detailed diary throughout the field study (Figure 2) and was generally positive about it, stating:

> *I liked the journal a lot. I particularly noticed I liked to fill in the exercises and questions, apart from writing down what I did that day. It made me realize what happens and who are important to me. Writing down my days made me realize how much I accomplish in a day.*

The diary provoked personal reflection about how caregiving made them feel and why and about patterns in their own behavior. For example, P7 stated that rereading how hectic her previous day was made her realize why she was so tired at the moment. This helped her to decide she should have a night off. Thus, participants also indicated that the journal made them see all the different things they had accomplished in a day. This supported appreciation of the nice moments they experienced. Especially, the themed pages in the journal were helpful. They helped seeing and appreciating what was important and to look at things with a different perspective, such as identifying small upsides in stressful situations. P7 indicated that the page about people involved in the care process (see Figure 3, top) helped her see that there were more people she could ask for help than she realized. P4 used the same pages in the journal to reflect on the quality in the communication with the people involved (Figure 3, bottom).

Heart Rate Measurements

Tracking physiological data as an indication of how participants were feeling when they wore them induced some interesting findings. Except for one, all participants found the idea of measuring their physiological well-being interesting, especially for measuring how they felt during specific moments in which they expected an abnormal heart rate (eg, during stressful times, during caregiving, or after exercise). For example, P5 stated,

> *In the beginning I thought about taking pictures during certain moments, especially in combination with the heart rate measurements. Because I thought it would be interesting to be able to see how it affected me.*

Figure 6. Left: P6 feels appreciated when she accompanies her grandma shopping. Right: P5 feels stressed and irritated when she took this photo: she had a lot of schoolwork to do but instead was on her way to help her parents.

Remarkably, the heart rate band got an extra meaning: it became a visual expression of being a caregiver. During the previous interviews, many caregivers explained they did not consciously think of themselves as *informal caregivers* and they often did not talk to their social network about their caregiver status. During the study, however, when a colleague or friends asked a participant about why he/she was wearing a heart rate band, it became a positive cue to talk about their caregiving situations with someone they would normally not.

> *I'm never very explicit about being a caregiver, but I noticed that when I was doing this study, that people automatically asked me about it. Because I was taking photos and wearing the heart rate wristband. Before, I always thought: Those are my private things. But people found it special that I was making photos and wearing the band. That made it easy to talk about it and people liked that I was doing it. They said: "Oh, I didn't even know that you were a caregiver". It felt good to be able to tell them, to be honest.* [P2]

For the final interview, the heart rate measurements that were recorded during the study were plotted on a daily timeline that also showed the photos that the participant had taken each day (see Figure 5). This combination proved helpful in remembering the context and details of a certain moment and helped put days with negative experiences into perspective as it showed other more positive experiences on the same day as well. For example, when discussing her photos, P5 picked out a photo where she stated that she was very stressed and irritated (Figure 6). However, when the same photo came up in the heart rate plot (Figure 5) she remembered that she ended that same day on a very positive note together with her brother. This was something she had forgotten all about.

There was 1 case with negative feedback in relation to the heart rate wristband. Owing to a technical error, the wrist band's battery indication light started blinking unexpectedly and this discouraged a participant (P2) to take any more measurements during the field study because it worried her that something was wrong with her heart rate. Overall, the visual representation of their well-being from the heart rate measurements was found reassuring by most of the participants. For example, P4 said,

> *I am relieved that the measurements don't show anything too surprising. Sometimes I'm afraid that it gets to me more than I think. It's reassuring.*

A Deeper Understanding of Caregivers' Experiences

Through the study, caregivers in our study became (more) aware of their experience of caregiving. A number of dynamic conditions seemed to affect the caregivers particularly in their emotional states and well-being:

- *How the care recipient is doing:* Caregivers expressed that their personal happiness often depended on how the person they cared for was doing and that they were frequently concerned about their loved one, especially when they were not easily able to check in. For example, 1 caregiver (P1) lived far away from his father, making it difficult for him to be there: "I worry a lot about my father, especially because he lives so far away." Moreover, participants indicated that moments of worry arose around critical moments for the care recipient, for example, during hospital appointments; when waiting on test results; or when hearing about bad outcomes that influence the treatment of the person they care for. This emotional dependence was also found in previous research [8] but the samples from the kit (eg, the photos) made this more visible.

- *Uncertainties and unawareness:* Especially before the field study, caregivers were often unaware of the time they spent caregiving and they were more focused on the health of the person they cared for than their own. After the field study, however, they were able to better indicate how their care responsibilities also affected them in moments that they were previously unaware (eg, feeling guilty about being at a sports game instead of with the care recipient). Moreover, having a visual overview (as with the journal entries or the heart rate visualizations) helped caregivers to see the bigger picture and reevaluate even negative experiences. Moreover, caregivers reported how capturing how they felt over time helped reassure them that they were doing well. It also made them more aware of their role. P6 stated: “I liked the exercises in between the journal entries. They make you think about how you experience caregiving, which things I like, about the future, etc. It makes you more conscious.”
- *Increased responsibility:* Older caregivers discussed how they had to adapt their work schedules, for example, by working less hours. Younger caregivers reported about falling behind on schoolwork or even obtaining severe study delay because of the extent of their care tasks. It was difficult for them to study at home when they experienced emotional burden as an effect of their loved one being ill. There was a clear need expressed for support on this level.
- *Talking about the care role and responsibilities to others:* In general, participants did not talk about their experiences and emotions regarding caregiving with friends much and they were more likely to talk about it with their partner or with other caregivers (if at all). The reasons mentioned for this were as follows: not feeling the need to talk about it with people to whom it did not concern or not wanting to bring it up because it was difficult to tell outsiders. P6 stated: “I just always think it’s hard to tell people I am a caregiver; it has such a negative connotation...” Motivations for caregivers to bring up their caregiving situation to their surroundings were more likely to be out of practical nature than social, such as informing their professional environment when they needed flexibility from work responsibilities. Interestingly, an unexpected outcome of using the heart rate band during the field study was that it seemed to produce a positive angle for caregivers to speak about their care role and responsibilities with people with whom they normally would not. P4 also added: “Participating in the study made me talk to my husband more about the situation.”

Evaluation of the Study Toolkit

Despite the fact that caregivers are a hard-to-reach user group for study participation because they tend to have little time, the tools in the Co-Care-KIT each lead to the recording of rich data of their experience of caregiving. In comparison with the interviews conducted before the field study, the Co-Care-KIT revealed a deeper understanding of the caregivers’ experiences. Much effort was put into learning about the personal multifaceted experiences during the introductory interviews. However, mainly the highlights of caregiving experiences were discussed. Although participants reported that the study was provoking at times, it also increased awareness of their own well-being and the full extent of their caregiving experience filled with stressors, as well as positive experiences. Caregivers also became more aware of the time spent caregiving. P1, for example, explained,

> *I learned how to name/recognize certain activities. From the kit I learned that I spend WAY more time caregiving than I thought.*

This insight could have lasting benefits, as P1 consequently stated: “I am going to push harder to get more formal help.” Especially because caregivers tend to prioritize the health and well-being of the person they care for over their own, this was an important outcome. Furthermore, capturing experiences contributed to an ongoing reflection and empowered participants to open up about their caregiver status to their social network.

Out of all the tools in the Co-Care-KIT, the most positive reactions concerned using the photo elicitation app. Participants found that taking a photo with the app was a convenient method for capturing experiences. Taking a photo and indicating how they felt through the app supported reflection about how they were doing several times a day. P6 stated,

> *It made me more aware of what I do. It makes you chew over certain moments: what I was doing or what I did.*

Several participants indicated that this small repetitive reflection had a positive effect during the study, for example, P3 stated: “It was reassuring that I realized that I feel very stable.” and P1 said,

> *Taking a photo makes for an evaluation moment. It makes clear that you are doing something, part of a process. That motivates me in tracking the other aspects of that process as well.*

Talking about the moments captured in photos with participants helped researchers better understand the personal experience of caregivers and revealed conflicting feelings that caregivers can experience. Similarly, the journal was also helpful in evoking personal reflection about how caregiving made participants feel.

A particular strength of using the toolkit seemed to lie in the richness of the collected samples and how these made it possible to discuss them. The kit allowed caregivers to record many different data samples during only a 2-week field study. Combining various sources of samples in a visual overview, such as the visualizations that combined heart rate measurements and photos on a daily timeline (see Figure 5), was a helpful prompt during the final interviews because of the contextual details it showed. Discussing these with participants helped in creating a bigger picture of the caregivers’ individual experiences. It helped caregivers put things into perspective, instead of focusing on one moment in particular. This was not only useful for gaining a deeper understanding of caregiver experiences for the researchers. Caregivers were positive about this aspect of the study as well, as it was also insightful to them. The idea of measuring their well-being and health was found interesting by the majority of participants, especially as a reassurance during specific moments that they suspected would affect them (eg, during stressful moments, during caregiving, or after exercise). Finally, the participants seemed to like the reflecting of the positive brought about by the kit, as P1 stated,

> *The kit helped to focus more on the positive, instead of thinking about the practical things you did that day (eg, cleaning), you remember the feeling you had better.*

Discussion

Principal Findings

The developments in the field of HCI, in which researchers (inspired by positive psychology) have begun to explore and define how technology can foster users' well-being [16,19,20], call for new approaches for understanding users' emotional experiences. This paper builds on those developments by exploring how technology can help capture users' changing emotional states *in situ* during a 2-week field study. Several aspects of the Co-Care-KIT study method particularly showed promise in supporting caregivers through (1) experience capturing, (2) collaborative reflection, (3) providing reassurance, and (4) social sharing:

1. *Support capturing the bigger (positive) picture:* Solely capturing experiences repeatedly throughout the field study gave participants an opportunity to reflect on their emotional experience (in accordance with Isaacs et al [35]). This ongoing reflection promoted a focus on the bigger picture, instead of only the (positive or negative) highlights of a situation. Furthermore, consistent with other research [50,52,64], the study found that writing, listing accomplishments, and reflecting on specific topics had positively affected participants' awareness of positive aspects in life. Of all the methods, participants generally enjoyed the photo experience sampling tool the most. This method took little effort and, thus, could provide interesting perspectives to deploy in further longer studies to capture and react to changes in wellness.
2. *Enabling collaborative reflection:* The Co-Care-KIT was found to be a valuable tool for discussing users' multifaceted experiences. The combination of allowing users to self-report and encouraging reflection resulted in a better articulation of their needs. Visualizing the collected data samples made it possible to then analyze experiences in collaboration with participants. The outcome of this collaborative reflection was not only relevant to the researchers but also an insightful process for participants. Mainly in the focus on this collaborative analysis, the Co-Care-KIT toolkit approach differs from similar research methods such as cultural probes.
3. *Justly reassure users of their health and well-being to increase their confidence in the ability to handle the situation:* There was early evidence that providing reassurance about the caregivers' health promoted confidence in their emotional ability to handle the situation. Given that many factors can affect the well-being of caregivers over time, this is an interesting finding that deserves further investigation. Overall, focusing on emphasizing and visualizing positive aspects of behavior can indeed promote self-confidence and mental resilience, which are important to caregivers [33].
4. *Provide tools for social sharing:* Wearing the heart rate band provoked discussion among caregivers with their social network about their caregiving status. Moreover, the purpose of being a part of the study empowered the caregivers to do so from a positive perspective. The study produced this unexpected positive side effect, as caregivers often tend to keep their experience to themselves. This highlights that providing empowering tools and positive cues can stimulate not only the awareness of the caregivers' emotional experience to caregivers, but also to others surrounding them.

Limitations and Future Work

Caregivers form a difficult-to-reach user group for study participation, as they are already heavily taxed. In concurrence with other studies [6,8], challenges came up in recruiting participants for the more extensive field study, resulting in a small sample size. Participants were easier to recruit in the initial questionnaire study [8] than for the more time-consuming study with the toolkit. The study also seemed to attract participants who were less burdened. The recruitment was most effective at including young caregivers as participants. It has to be noted that the sample is not fully representative for the group of informal caregivers, for example, in terms of male and female participants.

Furthermore, according to an extensive study [77], a large percentage of caregivers are interested in technology, but only few are using it. Thus, this still needs to be addressed in real-world settings. Besides, the Co-Care-KIT was used in a short field study. The aim was to investigate how different methods for capturing experiences each helped gain insight into the experience of caregivers. The themes identified in the applied thematic analysis of the qualitative data may have been affected by a researcher effect in which the data collected, and interpretations generated were partly informed by the investigators' perceptions [76].

Although it was beyond the scope of this study, it could be interesting to study the emotional states of caregivers over time by, for example, using one of the demonstrated tools for tracking experiences in a longer study (eg, long-term reflection on life events [35]). Furthermore, the field study lasted for 2 weeks, whereas caregivers often serve in a long-term situation. Thus, this study provides only a limited insight into changing experiences and in measuring the impact on emotional experiences and well-being. Even so, the method itself, particularly the journal and photo experience sampling tool, showed promise in giving caregivers greater awareness of their own well-being and encouraged social sharing and could become part of a solution. The study also showed an explicit need for more lightweight approaches and solutions that relieve the burden of caregivers. These should further respond to the changing emotional day-to-day experiences of caregivers in a positive way.

Conclusions

There is a vast opportunity for HCI to better understand and support informal caregivers' emotional wellness during caregiving through technology. This study showed the relevance of studying the emotional experience of informal caregivers and demonstrated a method for capturing those rich experiences *in situ* with a reflective toolkit. The toolkit was evaluated during a 2-week field study. This study found that through using the Co-Care-KIT method, participants became better able to articulate how their caregiving experiences influenced their well-being. The toolkit method enabled a collaborative analysis together with the participants about both the stressors (such as worrying, changes in their relationship with the care recipient, and having to adapt to the increased responsibilities that come with the caregiver role) as well as the positive experiences of caregivers (such as moments of gratitude, talking about caregiving with their social network, and finding reassurance in handling the situation). This study contributes to a better understanding of the day-to-day experiences of caregivers, which is a first step toward designing solutions for supporting the diverse and changing needs of caregivers at the right time and in the right context.

Acknowledgments

The authors would like to thank the study participants, Rick Kwekkeboom, the members of the Community Care research group, AHTI and the *Expertiselab voor jonge mantelzorgers*. This research is part of the project Co-Care-IT that has been funded by the AUAS program Urban Vitality.

Conflicts of Interest

None declared.

References

1. Broese van Groenou MI, De Boer A. Providing informal care in a changing society. Eur J Ageing 2016;13(3):271-279 [FREE Full text] [doi: 10.1007/s10433-016-0370-7] [Medline: 27610055]
2. Colombo F, Llena-Nozal A, Mercier J, Tjadens F. Help Wanted? Providing and Paying for Long-Term Care. Paris: OECD Publishing; 2011.
3. Won CW, Fitts SS, Favaro S, Olsen P, Phelan EA. Community-based "powerful tools" intervention enhances health of caregivers. Arch Gerontol Geriatr 2008;46(1):89-100 [FREE Full text] [doi: 10.1016/j.archger.2007.02.009] [Medline: 17467080]
4. Daniel K, Cason C, Ferrell S. Assistive technologies for use in the home to prolong independence. In: Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments. Rhodes, Greece: ACM Press; 2009 Presented at: PETRA'09; June 9, 2009; New York. [doi: 10.1145/1579114.1579140]
5. Lee M, Dey A. Using lifelogging to support recollection for people with episodic memory impairment and their caregivers. In: Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments Article No 14. New York: ACM Press; 2008 Presented at: HealthNet '08; June 17, 2008; Breckenridge, Colorado. [doi: 10.1145/1515747.1515765]
6. Robben S, Bosch L, Kanis M, Decancq J, Wiggers P. Managing flexible care with a context aware system for ageing-in-place. In: EAI Endorsed Transactions on Context-aware Systems and Applications. 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2015; Istanbul, Turkey p. 338-341. [doi: 10.4108/icst.pervasivehealth.2015.259229]
7. van Hoof J, Kort H, Rutten P, Duijnstee M. Ageing-in-place with the use of ambient intelligence technology: perspectives of older users. Int J Med Inform 2011;80(5):310-331. [doi: 10.1016/j.ijmedinf.2011.02.010] [Medline: 21439898]
8. Bosch L, Kanis M. Design opportunities for supporting informal caregivers. In: Extended Abstracts of CHI'16. New York: ACM Press; 2016 Presented at: CHI'16; May 07-12, 2016; San Jose, California, USA p. 2790-2797. [doi: 10.1145/2851581.2892354]
9. Cruijsberg M, Vastenburg M, Stappers PJ. Family carebook: a case study on designing peace of mind for family caregivers. In: Ambient Intelligence - Software and Applications Advances in Intelligent and Soft Computing. Berlin: Springer; 2012:129-136.
10. de Carvalho AFP, Tellioğlu H, Hensely-Schinkinger S, Habiger M. Supporting informal carers' independency through coordinated care. In: Proceedings of the 12th International Conference on the Design of Cooperative Systems (COOP '16). Switzerland: Springer International Publishing; 2016 Presented at: COOP'16; 23-27 May, 2016; Trento, Italy p. 155-172.
11. Magnusson L, Hanson E, Nolan M. The impact of information and communication technology on family carers of older people and professionals in Sweden. Ageing Soc 2005;25(5):693-713. [doi: 10.1017/S0144686X05003673]
12. Tixier M, Lewkowicz M. "Counting on the group" Reconciling online and offline social support among older informal caregivers. In: Proceedings of CHI'16. New York: ACM Press; 2016.

13. Zheng R, Daniel C, Hill R, Luptak M, Hicken B, Grant M. The effects of high and low technologies on emotional support for caregivers. International Journal of Cyber Behavior, Psychology and Learning 2013;3(3):34-45. [doi: 10.4018/ijcbpl.2013070103]
14. Adler R, Mehta R. Caregiving.org. Catalyzing technology to support family caregiving National Alliance for Caregiving; 2014 URL: https://www.caregiving.org/wp-content/uploads/2010/01/Catalyzing-Technology-to-Support-Family-Caregiving_FINAL.pdf [accessed 2019-02-12] [WebCite Cache ID 768G7zqtG]
15. Chen Y, Ngo V, Park S. Caring for caregivers: Designing for integrality. Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13); New York: ACM Press; 2013 Presented at: CSCW'13; February 23-27, 2013; San Antonio, Texas, USA p. 91-102.
16. Kanis M, Brinkman WP, Perry M. Designing for positive disclosure: what do you like today? Int J Ind Ergon 2009;39(3):564-572. [doi: 10.1016/j.ergon.2008.10.001]
17. Kanis M, Perry M, Brinkman WP. Positive expressive technologies for mental wellness. In: Proceedings of CHI'08 workshop on Technology in Mental Health. 2008 Presented at: CHI'08 workshop on Technology in Mental Health; 2008; Florence, Italy.
18. Peters D, Calvo R. Positive Computing: Technology for Wellbeing and Human Potential. Cambridge, Massachusetts: MIT Press; 2014.
19. Riva G, Baños RM, Botella C, Wiederhold BK, Gaggioli A. Positive technology: using interactive technologies to promote positive functioning. Cyberpsychol Behav Soc Netw 2012;15(2):69-77. [doi: 10.1089/cyber.2011.0139]
20. Sander T. Positive Computing. In: Biswas-Diener R, editor. Positive psychology as social change. The Netherlands: Springer; 2011:309-326.
21. Dubberly H, Mehta R, Evenson S, Pangaro P. Reframing health to embrace design of our own well-being. Interactions 2010 May 01;17(3):56-63. [doi: 10.1145/1744161.1744175]
22. Fitzpatrick G, Ellingsen G. A review of 25 years of CSCW research in healthcare: Contributions, challenges and future agendas. Comput Support Coop Work 2012;22(4-6):609-665. [doi: 10.1007/s10606-012-9168-0]
23. Bastawrous M. Caregiver burden--a critical discussion. Int J Nurs Stud 2013 Mar;50(3):431-441. [doi: 10.1016/j.ijnurstu.2012.10.005] [Medline: 23131724]
24. Harding R, Gao W, Jackson D, Pearson C, Murray J, Higginson IJ. Comparative analysis of informal caregiver burden in advanced cancer, dementia, and acquired brain injury. Journal of Pain and Symptom Management 2015;50(4):445-452. [doi: 10.1016/j.jpainsymman.2015.04.005]
25. Hogstel M, Cox CL, Walker C. Caring for older adults: the benefits of informal family caregiving. J Theory Constr Test 2005;9(2):55-61 [FREE Full text]
26. van der Lee J, Bakker TJ, Duivenvoorden HJ, Dröes RM. Multivariate models of subjective caregiver burden in dementia: a systematic review. Ageing Res Rev 2014 May;15:76-93. [doi: 10.1016/j.arr.2014.03.003] [Medline: 24675045]
27. Razali R, Ahmad F, Rahman FN, Midin M, Sidi H. Burden of care among caregivers of patients with Parkinson disease: a cross-sectional study. Clin Neurol Neurosurg 2011 Oct;113(8):639-643. [doi: 10.1016/j.clineuro.2011.05.008] [Medline: 21684679]
28. Savage S, Bailey S. The impact of caring on caregivers' mental health: a review of the literature. Aust Health Rev 2004;27(1):111-117. [doi: 10.1071/AH042710111] [Medline: 15362303]
29. Kramer BJ. Gain in the caregiving experience: where are we? What next? Gerontologist 1997 Apr;37(2):218-232. [doi: 10.1093/geront/37.2.218] [Medline: 9127978]
30. Cohen CA, Colantonio A, Vernich L. Positive aspects of caregiving: rounding out the caregiver experience. Int J Geriatr Psychiatry 2002 Feb;17(2):184-188. [Medline: 11813283]
31. Lloyd J, Patterson T, Muers J. The positive aspects of caregiving in dementia: a critical review of the qualitative literature. Dementia (London) 2016 Nov;15(6):1534-1561. [doi: 10.1177/1471301214564792] [Medline: 25547208]
32. Peacock S, Forbes D, Markle-Reid M, Hawranik P, Morgan D, Jansen L, et al. The positive aspects of the caregiving journey with dementia: using a strengths-based perspective to reveal opportunities. J Appl Gerontol 2009 Aug 25;29(5):640-659. [doi: 10.1177/0733464809341471]
33. Braithwaite V. Contextual or general stress outcomes: making choices through caregiving appraisals. Gerontologist 2000 Dec;40(6):706-717. [doi: 10.1093/geront/40.6.706] [Medline: 11131087]
34. Acton GJ. Health-promoting self-care in family caregivers. West J Nurs Res 2002 Feb;24(1):73-86. [doi: 10.1177/01939450222045716] [Medline: 11829276]
35. Isaacs E, Konrad A, Walendowski A, Lennig T, Hollis V, Whittaker S. Echoes from the past: how technology mediated reflection improves well-being. In: Proceedings of CHI'13. New York: ACM Press; 2013 Presented at: CHI'13; April 27-May 02, 2013; Paris, France p. 1071-1080. [doi: 10.1145/2470654.2466137]
36. Mayo Clinic. iTunes. 2012. AnxietyCoach URL: https://itunes.apple.com/us/app/anxietycoach/id565943257 [accessed 2019-02-12] [WebCite Cache ID 768GLX6eg]
37. Dickerson R, Gorlin E, Stankovic J. Empath: a continuous remote emotional health monitoring system for depressive illness. New York: ACM Press; 2011 Presented at: Proceedings of Wireless Health - WH'11; October 10-13, 2011; San Diego, California URL: https://www.cs.virginia.edu/~stankovic/psfiles/robempath.pdf

38. Kocielnik R, Sidorova N, Maggi F, Ouwerkerk M, Westerink J. Smart technologies for long-term stress monitoring at work. New Jersey: IEEE; 2013 Presented at: The 26th IEEE International Symposium on Computer-Based Medical Systems; June 20-22, 2013; Porto, Portugal p. 53-58. [doi: 10.1109/CBMS.2013.6627764]
39. Robben S, Aicha A, Kröse B. Measuring regularity in daily behavior for the purpose of detecting Alzheimer. In: Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2016 Presented at: Pervasive health’16; May 16-19, 2016; Cancun, Mexico p. 97-100.
40. Wang R, Aung M, Abdullah S, Brian R, Campbell A, Choudhury T. CrossCheck: Toward passive sensing and detection of mental health changes in people with schizophrenia. In: Proceedings of UbiComp'16. New York: ACM Press; 2016 Presented at: UbiComp'16; 2016; Heidelberg, Germany p. 886-897 URL: http://pac.cs.cornell.edu/pubs/Ubicomp2016_Crosscheck.pdf
41. Sellen AJ, Whittaker S. Beyond total capture: a constructive critique of lifelogging. Commun ACM 2010;53(5):70-77. [doi: 10.1145/1735223.1735243]
42. Bryant FB, Smart CM, King SP. Using the past to enhance the present: boosting happiness through positive reminiscence. J Happiness Stud 2005;6(3):227-260. [doi: 10.1007/s10902-005-3889-4]
43. Immordino-Yang MH, Christodoulou JA, Singh V. Rest is not idleness: implications of the brain's default mode for human development and education. Perspect Psychol Sci 2012 Jul;7(4):352-364. [doi: 10.1177/1745691612447308] [Medline: 26168472]
44. Cheng J, Bapat A, Thomas G, Tse K, Nawathe N, Crockett J, et al. GoSlow: Designing for slowness, reflection and solitude. In: Extended Abstracts of CHI’11. New York: ACM Press; 2011:429-438.
45. Hallnäs L, Redström J. Slow technology–Designing for reflection. Personal Ub Comp 2001;5(3):201-212. [doi: 10.1007/PL00000019]
46. Konrad A, Isaacs E, Whittaker S. Technology-mediated memory: is technology altering our memories and interfering with well-being? ACM Trans Comput-Hum Interact 2016;23(4). [doi: 10.1145/2934667]
47. Li I, Dey A, Forlizzi J. Understanding my data, myself: supporting self-reflection with Ubicomp technologies. In: Proceedings of the 13th international conference on Ubiquitous computing. New York: ACM Press; 2011 Presented at: Ubicomp'11; September 17-21, 2011; Beijing, China p. 405-414. [doi: 10.1145/2030112.2030166]
48. Prilla M, Degeling M, Herrmann T. Collaborative reflection at work: supporting informal learning at a healthcare workplace. In: Proceedings of the 17th ACM international conference on Supporting group work.: ACM Press; 2012 Presented at: The 17th ACM international conference on Supporting group work; October 27-31, 2012; Sanibel Island, Florida, USA p. 55-64. [doi: 10.1145/2389176.2389185]
49. Carter S, Mankoff J. When participants do the capturing: the role of media in diary studies. In: Proceedings of CHI'05. New York: ACM Press; 2005 Presented at: CHI'05; April 02-07, 2005; Portland, Oregon, USA p. 899-908. [doi: 10.1145/1054972.1055098]
50. Burton C, King L. The health benefits of writing about intensely positive experiences. J Res Pers 2004 Apr;38(2):150-163. [doi: 10.1016/S0092-6566(03)00058-8]
51. Sexton JD, Pennebaker JW, Holzmueller CG, Wu AW, Berenholtz SM, Swoboda SM, et al. Care for the caregiver: benefits of expressive writing for nurses in the United States. Prog Palliat Care 2009;17(6):307-312. [doi: 10.1179/096992609X12455871937620]
52. Smyth JM, Pennebaker JW. Exploring the boundary conditions of expressive writing: In search of the right recipe. Br J Health Psychol 2008 Feb;13(1):1-7. [doi: 10.1348/135910707X260117] [Medline: 18034917]
53. Gaver B, Dunne T, Pacenti E. Design: cultural probes. interactions 1999;6(1):21-29. [doi: 10.1145/291224.291235]
54. Schorch M, Wan L, Randall D, Wulf V. Designing for those who are overlooked - Insider perspectives on care practices and cooperative work of elderly informal caregiver. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York: ACM Press; 2016 Presented at: CSCW'16; February 27-March 02, 2016; San Francisco, California, USA p. 785-797. [doi: 10.1145/2818048.2819999]
55. Sanders EB, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
56. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. Annu Rev Psychol 2003;54:579-616. [doi: 10.1146/annurev.psych.54.101601.145030] [Medline: 12499517]
57. Hektner J, Schmidt J, Csikszentmihalyi M. Experience Sampling Method: Measuring the Quality of Everyday Life. Thousand Oaks, California: SAGE Publications; 2006.
58. Lyubomirsky S, Kurtz J. Positive Psychology. In: Handbook of Research Methods for Studying Daily Life. New York: Guilford Publications; 2013:553-569.
59. Hicks J, Ramanathan N, Kim D, Monibi M, Selsky J, Hansen M, et al. AndWellness: an open mobile system for activity and experience sampling. Wireless Health '10 2010:34-43 [FREE Full text]
60. Weppner J, Serino S, Cipresso P, Lukowicz P, Gaggioli A, Riva G. Smartphone based experience sampling of stress-related events. New York: ACM Press; 2013 Presented at: Proceedings of Pervasive Health’13; 2013; Venice, Italy p. 464-467. [doi: 10.4108/icst.pervasivehealth.2013.252358]

61. Yue Z, Litt E, Cai C, Stern J, Baxter K, Guan Z, et al. Photographing information needs: the role of photos in experience sampling method-style research. In: Proceedings of CHI'14. New York: ACM Press; 2014 Presented at: CHI'14; 2014; Toronto, Canada p. 1545-1554.
62. Plunkett R, Leipert BD, Ray SL. Unspoken phenomena: using the photovoice method to enrich phenomenological inquiry. Nurs Inq 2013 Jun;20(2):156-164. [doi: 10.1111/j.1440-1800.2012.00594.x] [Medline: 22381071]
63. Guerra S, Rodrigues S, Demain S, Figueiredo D, Sousa L. Evaluating proFamilies-dementia: adopting photovoice to capture clinical significance. Dementia (London) 2013 Sep;12(5):569-587. [doi: 10.1177/1471301212437779] [Medline: 24337330]
64. Baumer E. Reflective informatics: Conceptual dimensions for designing technologies of reflection. New York: ACM Press; 2015 Presented at: CHI '15; 2015; Seoul, Republic of Korea p. 585-594. [doi: 10.1145/2702123.2702234]
65. Mio Global. 2016. URL: http://www.mioglobal.com [accessed 2019-02-12] [WebCite Cache ID 768FUaZ0p]
66. Marco Altini. 2013. Heart Rate Variability Logger URL: https://www.marcoaltini.com/blog/heart-rate-variability-logger-app-details [accessed 2019-02-12] [WebCite Cache ID 768Ff4vtT]
67. Boonnithi S, Phongsuphap S. Comparison of heart rate variability measures for mental stress detection. Comput Cardiol (2010) 2011;38:85-88.
68. Healey J. Physiological sensing of emotion. In: Calvo R, D'Mello S, Gratch J, Kappas A, editors. The Oxford Handbook of Affective Computing. Oxford: Oxford University Press; 2014.
69. Nguyen TA, Zeng Y. A physiological study of relationship between designer's mental effort and mental stress during conceptual design. Comput Aided Des 2014 Sep;54:3-18. [doi: 10.1016/j.cad.2013.10.002]
70. Schubert C, Lambertz M, Nelesen RA, Bardwell W, Choi J, Dimsdale JE. Effects of stress on heart rate complexity--a comparison between short-term and chronic stress. Biol Psychol 2009;80(3):325-332 [FREE Full text] [doi: 10.1016/j.biopsycho.2008.11.005] [Medline: 19100813]
71. McDarby G. Mind Games group. 2005. URL: http://medialabeurope.org/research/group.php%3Fid=7.html [accessed 2019-04-30] [WebCite Cache ID 781wGMalb]
72. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
73. Smyth C, Blaxland M, Cass B. 'So that's how I found out I was a young carer and that I actually had been a carer most of my life'. Identifying and supporting hidden young carers. J Youth Stud 2011;14(2):145-160. [doi: 10.1080/13676261.2010.506524]
74. Watkins J, Stanton L, Saunders B, Lasocki G, Chung P, Hibberd P. Working in partnership with family carers: the importance of learning from carers' experiences. Qual Ageing Older Adults 2011;12(2):103-108. [doi: 10.1108/14717791111144713]
75. Robinson BC. Validation of a Caregiver Strain Index. J Gerontol 1983;38(3):344-348. [doi: 10.1093/geronj/38.3.344]
76. Guest G, MacQueen K, Namey E. Applied Thematic Analysis. Los Angeles, USA: Sage Publications; 2011.
77. AARP. 2016. Caregivers & technology: What they want and need URL: https://www.aarp.org/content/dam/aarp/home-and-family/personal-technology/2016/04/Caregivers-and-Technology-AARP.pdf [accessed 2019-05-01] [WebCite Cache ID 781xb4RDU]

Abbreviations

CSI: Caregiver Strain Index
ESM: experience sampling method
HCI: human-computer interaction
HRV: heart rate variability

Edited by J Torous; submitted 12.02.19; peer-reviewed by M Perry, G Baumblatt, N Heynsbergh; comments to author 20.03.19; revised version received 03.04.19; accepted 04.04.19; published 28.05.19.

Please cite as:
Bosch L, Kanis M, Dunn J, Stewart KA, Kröse B
How Is the Caregiver Doing? Capturing Caregivers' Experiences With a Reflective Toolkit
JMIR Ment Health 2019;6(5):e13688
URL: http://mental.jmir.org/2019/5/e13688/
doi:10.2196/13688
PMID:31140435

Original Paper

Adverse Childhood Experiences Ontology for Mental Health Surveillance, Research, and Evaluation: Advanced Knowledge Representation and Semantic Web Techniques

Jon Hael Brenas[1], PhD; Eun Kyong Shin[1], PhD; Arash Shaban-Nejad[1], MPH, PhD

Department of Pediatrics, Oak Ridge National Laboratory Center for Biomedical Informatics, University of Tennessee Health Science Center, Memphis, TN, United States

Corresponding Author:
Arash Shaban-Nejad, MPH, PhD
Department of Pediatrics
Oak Ridge National Laboratory Center for Biomedical Informatics
University of Tennessee Health Science Center
50 N Dunlap Street, R492
Memphis, TN, 38103
United States
Phone: 1 901 287 5836
Email: ashabann@uthsc.edu

Abstract

Background: Adverse Childhood Experiences (ACEs), a set of negative events and processes that a person might encounter during childhood and adolescence, have been proven to be linked to increased risks of a multitude of negative health outcomes and conditions when children reach adulthood and beyond.

Objective: To better understand the relationship between ACEs and their relevant risk factors with associated health outcomes and to eventually design and implement preventive interventions, access to an integrated coherent dataset is needed. Therefore, we implemented a formal ontology as a resource to allow the mental health community to facilitate data integration and knowledge modeling and to improve ACEs' surveillance and research.

Methods: We use advanced knowledge representation and semantic Web tools and techniques to implement the ontology. The current implementation of the ontology is expressed in the description logic ALCRIQ(D), a sublogic of Web Ontology Language (OWL 2).

Results: The ACEs Ontology has been implemented and made available to the mental health community and the public via the BioPortal repository. Moreover, multiple use-case scenarios have been introduced to showcase and evaluate the usability of the ontology in action. The ontology was created to be used by major actors in the ACEs community with different applications, from the diagnosis of individuals and predicting potential negative outcomes that they might encounter to the prevention of ACEs in a population and designing interventions and policies.

Conclusions: The ACEs Ontology provides a uniform and reusable semantic network and an integrated knowledge structure for mental health practitioners and researchers to improve ACEs' surveillance and evaluation.

(JMIR Ment Health 2019;6(5):e13498) doi:10.2196/13498

KEYWORDS

ontologies; mental health surveillance; adverse childhood experiences; semantics; computational psychiatry

Introduction

The study of Adverse Childhood Experiences (ACEs) and their consequences in terms of diseases and health risks has emerged during the past 20 years [1]. The US Centers for Disease Control and Prevention (CDC) demonstrated that ACEs are one of the root causes of several physical, social, cognitive, and emotional concerns [2], and affect nearly half of all US children under 18 years of age [3]. Direct indicators, including emotional, physical, and sexual abuse, at the individual level and indirect home environment indicators are commonly considered when studying ACEs. In general, ACEs are measured by two main experiences during upbringing and within the home environment: victimization experience and exposure to adversities in the family. Victimization includes emotional abuse, physical abuse, and sexual abuse. The exposure to adversities, often referred to as household dysfunction, is measured by family members' history of substance abuse, mental illness, domestic violence, incarceration, marital status, and financial difficulties leading to food deficiency. However, these factors influence health outcomes with different degrees and are closely linked to other adversities at the community level and other social determinants of health (SDH). Over the years, a connection between a higher number of ACEs and a variety of negative outcomes (eg, substance abuse [3], impaired memory [4], biological systems impairments [5], or criminal activities [6]) has been identified. ACEs are known to have complex and negative influences on individual health outcomes throughout the life course and have enduring negative influences on physiological outcomes as well as psychological functioning. ACEs do not only affect individuals but also have intergenerational impacts. Yet, many questions related to the ACEs' causal pathway and the effectiveness of existing interventions have remained unanswered: for example, deciding on whether the best intervention for a particular scenario is medical or a community-based approach [7].

The complexity of ACEs' causal pathway demands comprehensive and multidimensional data coordination. To design effective preventive interventions to reduce the burdens of ACEs, researchers and clinicians need access to a consistent knowledge-driven, evidence-based comprehensive analytic framework to study and monitor the causes of ACEs and their impacts on health (ie, obesity, mental health, and substance abuse), education (ie, cognitive developments, educational attainment, and graduation rates), and social dimensions (ie, placement in foster care and involvement with the justice system). The ideal ACEs knowledge-based system should be able to capture and identify ACEs indicators, detect individuals and groups at high risk of ACEs, integrate and validate ACEs health and social determinants, as well as exposure and genetic variations at individual and population levels. In order to improve the surveillance of ACEs, it is crucial to access a standard vocabulary that facilitates data collection, analysis, interpretation, and exchange of data between different parties and disciplines. An ontology is a standardized computational artifact that is used to capture, represent, and reason about the knowledge in the field. Ontologies capture knowledge by defining concepts, instances, relationships, and axioms. They increase the interoperability between different data sources and systems and improve the dissemination of data and knowledge across different disciplines. They also allow the use of semantic technologies to reveal new associations between the datasets and, therefore, discover new knowledge through logical inference. Ontologies are widely used in health and biomedicine and have made a substantial contribution to translational and clinical research as well as public and personalized health care [8].

The creation of the ACEs Ontology is rooted in recent efforts to study the effects of early experiences and social-environmental factors on children's development and life-course health. These efforts have been made by multiple actors from different disciplines at various levels, including international organizations (eg, the World Health Organization [9]), nongovernmental organizations (eg, the Foundation for Excellence in Mental Health Care [10] and the Center for Youth Wellness [11]), government offices (eg, the Tennessee Department of Children's Services [12] and the CDC [13]), volunteers that consolidate their efforts (eg, the ACEs Connection community [14] and the Cumbria Resilience Project [15]), and academic studies. All of these projects have a common goal, which is preventing occurrences of ACEs and providing assistance to victims; however, their results, studies, interventions, and tools are limited by the fact that they cannot easily be shared in a way that would improve the ability to repeat or reuse them elsewhere. The ACEs Ontology can bridge the gap between these efforts for efficient ACEs studies and surveillance.

While several ontologies exist in the domain of mental health, the ACEs Ontology is designed to deal with a carefully restricted scope, which is the study, prevention, and treatment of ACEs. We have included in the ontology some aspects that we deem important because they are tangentially connected to ACEs, such as potential health outcomes as well as their possible causes and aggravating conditions. Many other projects and resources exist that provide more detailed views of these adjacent domains; for example, the Children's Health Exposure Analysis Resource (CHEAR) project [16] deals with environment as well as measurable exposure and biological responses. Since environmental factors have impacts on the development of children and on the incidence of ACEs, we plan to use existing resources, such as the CHEAR ontology [17], in future versions of the ACEs Ontology. The same can be said of other adjacent domains, including the legal and penal system, psychiatry and psychology of children and adults, nutrition, and interventions.

In this paper, we present the ACEs Ontology as a formal reusable resource that can be used by the mental health research community to advance the surveillance and study of ACEs. The ACEs ontological structure provides a semantic backbone supporting the entire data model and hierarchy and supports logical inference and query answering. The ontology is used to form a consensus on the kind of data that is relevant and to create a common lexicon that makes it easier to share and reuse knowledge and information in the domain. The ACEs Ontology is an integral part of the Semantic Platform for Adverse Childhood Experiences Surveillance (SPACES) [18], which is a semantic recommender system currently under development. The SPACES system employs the contextual knowledge provided by the ACEs Ontology, along with a hybrid content- and context-based filtering approach to enable intelligent exploratory and explanatory analysis and informed decision making. In this paper, we will first explain how the ontology was built and how it will dynamically expand to meet the needs of the ever-evolving field of ACEs' surveillance and prevention. We will then show the usability of the ontology through a set of use-case scenarios. The paper will conclude with discussions of the findings and suggestions for future directions.

Methods

To facilitate the integration of ACEs-related data from different data sources, the ontology reuses several existing ontological concepts and relations as much as possible, while introducing new components. In some cases, the ontology integrates multiple existing elements in order to represent new pieces of knowledge. The surveillance of ACEs is focused on monitoring, detection, and prevention of ACEs as well as studying their causality pathways and short- or long-term consequences. The knowledge required to describe these activities is linked to concepts and terminologies coming from a wide variety of domains, including medicine, legal justice, cognitive and personal behavior, and community responses. This means that building an integrated ontology and knowledge base that covers the whole range of relevant subjects is a tedious, time-consuming, and complex task.

To build the foundation for the ontology, we identified five different key pieces of knowledge in our scope of study. As shown in Figure 1, *Person* is used to representing human subjects. It encompasses the child as well as their family members and the rest of their social network. Felitti et al [1] provided a set of definitions and criteria for ACEs. ACEs may affect *Persons* at different levels of granularity (eg, divorce or separation of parents might affect all the children in a family, while being the victim of physical or sexual abuse might only affect one of them). *Social Determinants of Health* are upstream factors that can affect the *Person* and the exposure and manifestation of ACEs. SDH also act at different levels of granularity (eg, only members of a specific household are affected by the presence of mold at home, while the whole neighborhood is influenced, one way or another, by the lack of public transportation or safety). Proximity to a clinic or grocery store can be beneficial for all members of the community, while proximity to other institutions can be positive for only a portion of the community (eg, the presence of a strong religious community might positively influence its well-standing members, but may be seen negatively by people who do not belong). *Negative Health Outcomes* are the medical and surgical adversaries that can be caused or worsened by the ACEs (eg, higher risk of heart attack or cancer). *Intervention* represents the way that social workers, medical practitioners, or policy makers can mitigate or negate negative health effects. Interventions can be implemented through various channels (eg, medical, surgical, and legal). For instance, interventions can include heart surgery to treat a stroke or providing legal advice for people who fear that they might be evicted from their home. The ACEs Ontology contains more-detailed concepts to accurately describe these different overarching topics.

Figure 1. The top conceptual model for the Adverse Childhood Experiences (ACEs) Ontology, demonstrating the interactions between five major concepts: Person, ACEs, Social Determinants of Health, Interventions, and Negative Health Outcomes.

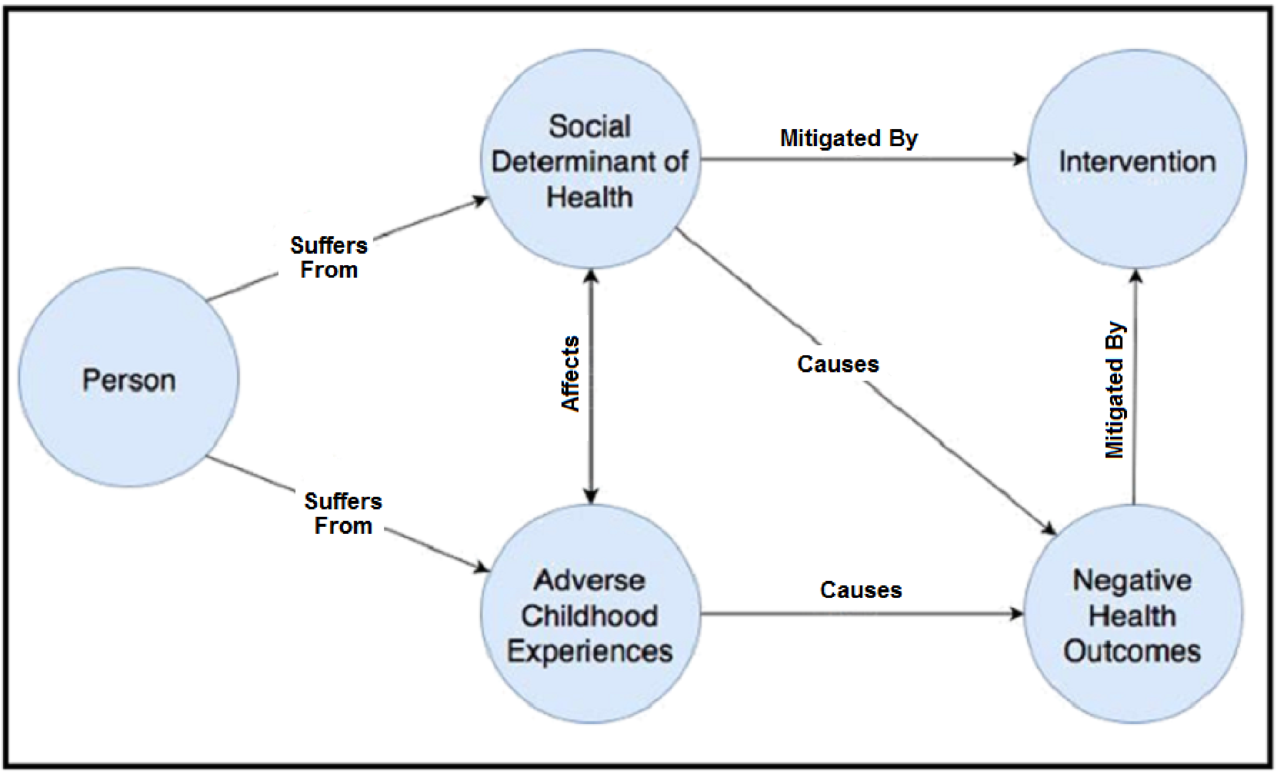


Figure 2. An abstract representation of some of the ontological concepts along with their relationships with the other thesauri in the field. SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms; ACEs: Adverse Childhood Experiences; NCIT: National Cancer Institute Thesaurus; MedDRA: Medical Dictionary for Regulatory Activities.

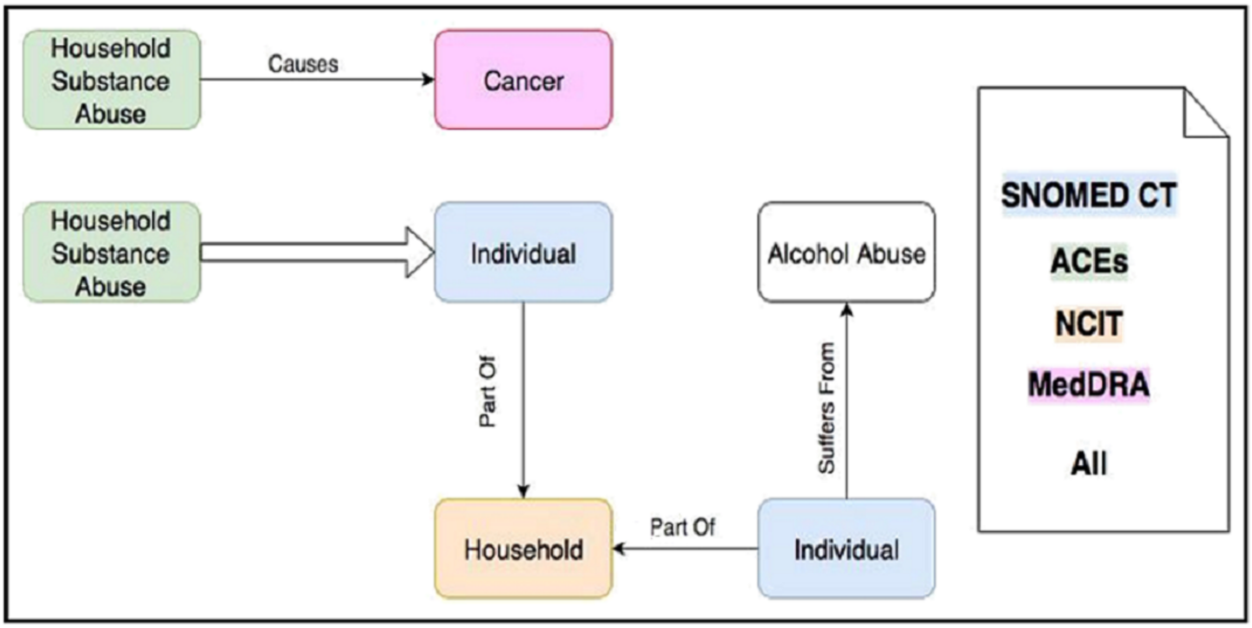


Figure 2 represents some examples of the major ontological elements; for instance, the ACE *Household Substance Abuse* affects *Individuals* who are part of the same *Household* as *Individuals* suffering from *Alcohol Abuse*— that *Household Substance Abuse* may itself be a causal factor for the development of *Cancer*. The concepts and relationships that appear in the ontology come from various sources in the field (eg, literature, databases, other ontologies, and domain experts). Many of the concepts that are central to ACEs (eg, abuse, violence, and mental illnesses and their medical consequences such as alcoholism, depression, obesity, and cancer) have been scattered among different existing biomedical ontologies and controlled vocabularies. For example, based on different needs, we imported a subset of the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) [19], the National Cancer Institute Thesaurus (NCIT) [20], and the Medical Dictionary for Regulatory Activities (MedDRA) [21] into the ACEs Ontology [22]. There is considerable overlap between the domains represented in these ontologies and, thus, it is not a surprise to find some repetitive concepts with similar or different definitions. In order to improve interoperability and reusability, we imported all the definitions and stated that they were equivalent, where applicable (ie, they describe the same concept).

Results

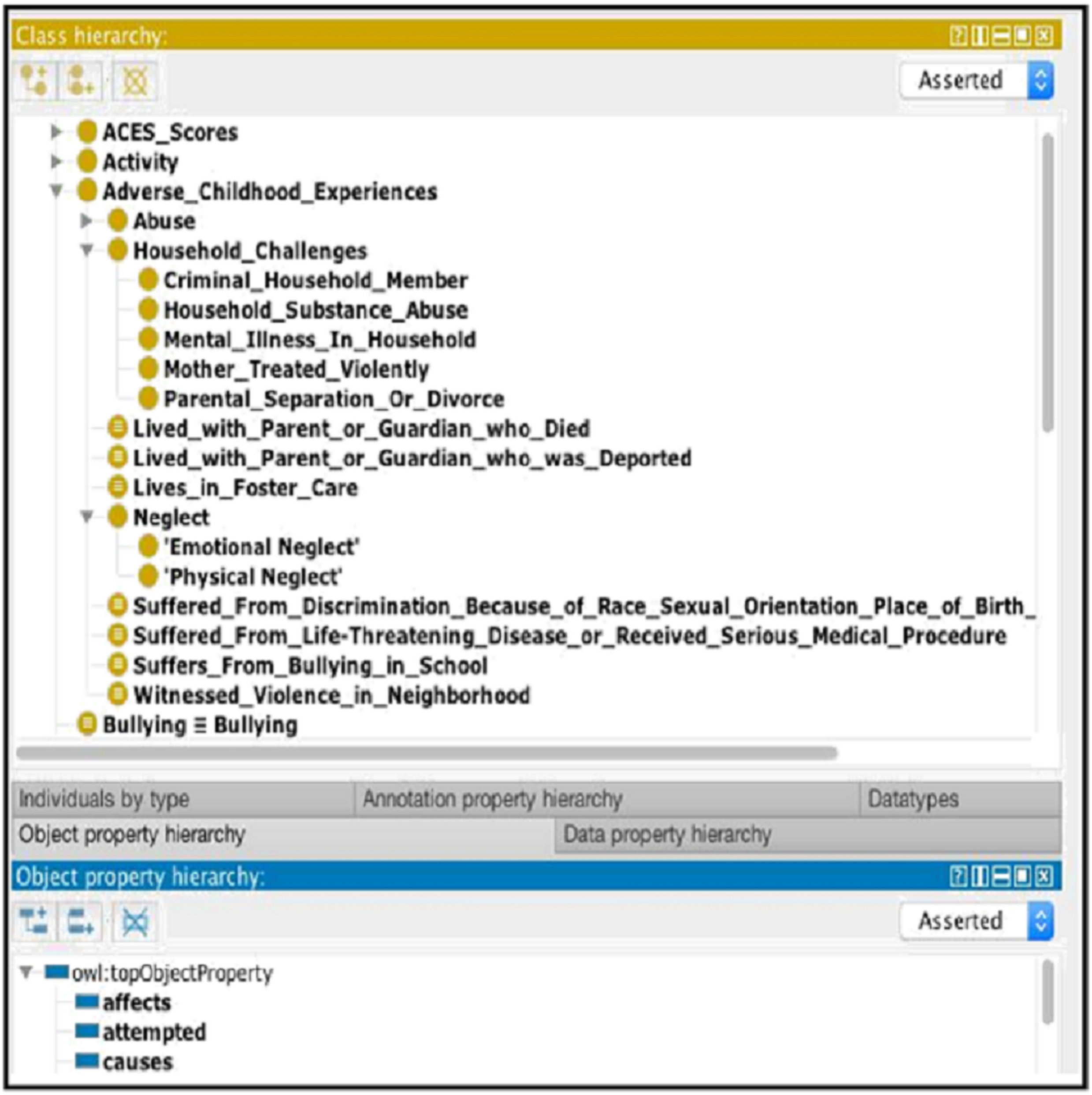
Overview

Figure 3 shows a fragment of the ACEs Ontology, including the partial concept and property hierarchies. The Web Ontology Language (OWL) format [23] is the de facto standard to express ontologies; for this reason, and for our ACEs Ontology to be freely reused by the community, the OWL format [23] of the ontology was made available via BioPortal [22] (see Figure 4).

An important decision when building the ontology was determining its overall expressivity. The more expressive the ontology, the higher the complexity of its associated reasoning tasks. In a similar way, increasing the complexity of an ontology makes it harder to maintain, as it becomes more difficult to make sure that no two inconsistent statements cohabit within. For this reason, and to encourage reusability of the ontologies, many biomedical ontologies focus almost exclusively on the taxonomical part of the ontology, which is defining inclusion relationships between various concepts (eg, $Verbal\ abuse \subseteq Abuse$; for instance, verbal abuse is a subcategory of abuse) and avoiding the definition of more complex axioms (eg, $Verbal\ abuse \cong Abuse \wedge \exists$ has Component. *Verbal*, for instance, verbal abuse is exactly the category of abuse that has a verbal component). In our ontology, we gradually define complex axioms as needed to capture more sophisticated key concepts necessary for ACEs' surveillance. General axioms, such as the ones shown in Figure 5, are used to express these complex concepts.

The current implementation of the ontology [22] is expressed in the description logic ALCRIQ (D), a sublogic of OWL 2 [23]. It is an expressive logic and, thus, the complexity of the decision problems is relatively high. In order to reduce the complexity, it is possible to avoid the use of datatypes, which are currently only used for dates and frequencies that might be handled as concepts; counting quantifiers, used to define the ACEs scores; or simplifying the role hierarchies. The ACEs Ontology currently contains 297 classes, 93 object properties, and 3 data properties. Most existing axioms are in the forms of class subsumption or equivalence.

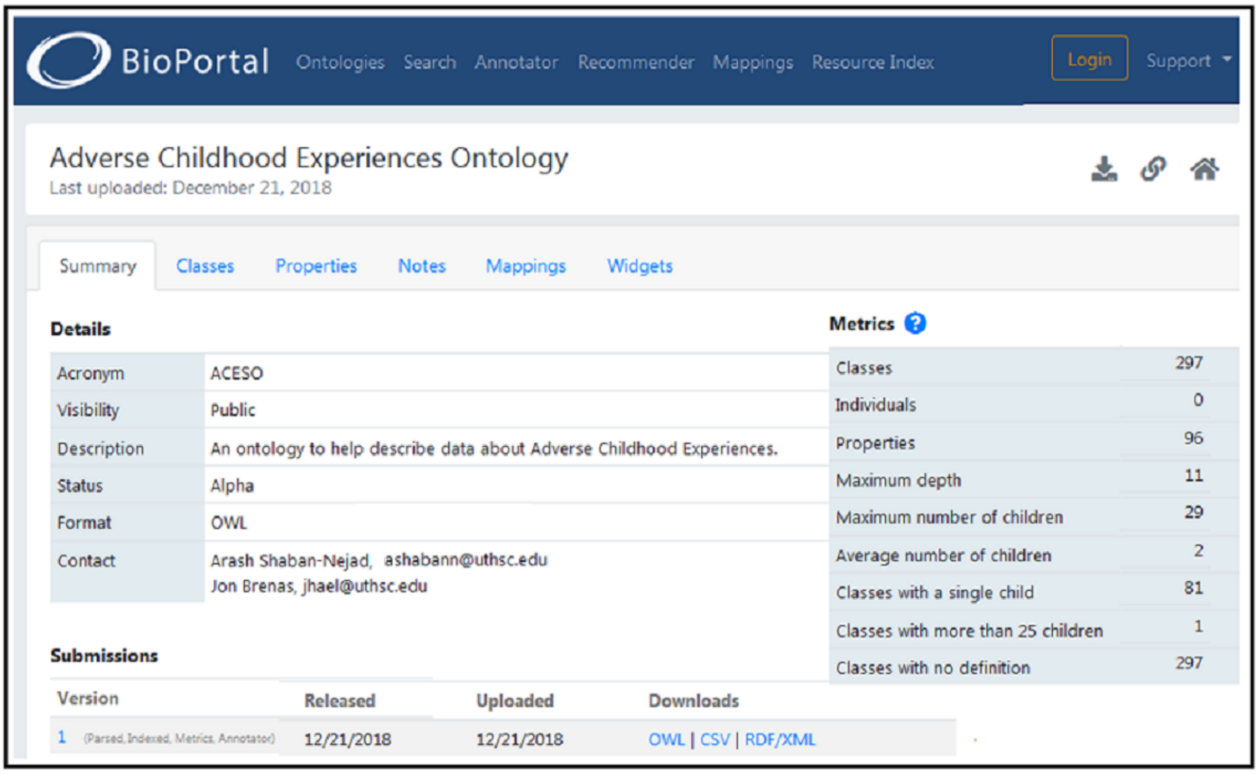
Figure 3. A fragment of the Adverse Childhood Experiences (ACEs) Ontology representing partial class and property hierarchies. OWL: Web Ontology Language.



- ACES_Scores
- Activity
- Adverse_Childhood_Experiences
  - Abuse
  - Household_Challenges
    - Criminal_Household_Member
    - Household_Substance_Abuse
    - Mental_Illness_In_Household
    - Mother_Treated_Violently
    - Parental_Separation_Or_Divorce
  - Lived_with_Parent_or_Guardian_who_Died
  - Lived_with_Parent_or_Guardian_who_was_Deported
  - Lives_in_Foster_Care
  - Neglect
    - 'Emotional Neglect'
    - 'Physical Neglect'
  - Suffered_From_Discrimination_Because_of_Race_Sexual_Orientation_Place_of_Birth_
  - Suffered_From_Life-Threatening_Disease_or_Received_Serious_Medical_Procedure
  - Suffers_From_Bullying_in_School
  - Witnessed_Violence_in_Neighborhood
- Bullying ≡ Bullying

- owl:topObjectProperty
  - affects
  - attempted
  - causes

Figure 4. The Adverse Childhood Experiences (ACEs) Ontology on the National Center for Biomedical Ontology's (NCBO) BioPortal [22]. ACESO: Adverse Childhood Experiences Ontology; OWL: Web Ontology Language.



Adverse Childhood Experiences Ontology

Last uploaded: December 21, 2018

Details

Acronym	ACESO
Visibility	Public
Description	An ontology to help describe data about Adverse Childhood Experiences.
Status	Alpha
Format	OWL
Contact	Arash Shaban-Nejad, ashabann@uthsc.edu Jon Brenas, jhael@uthsc.edu

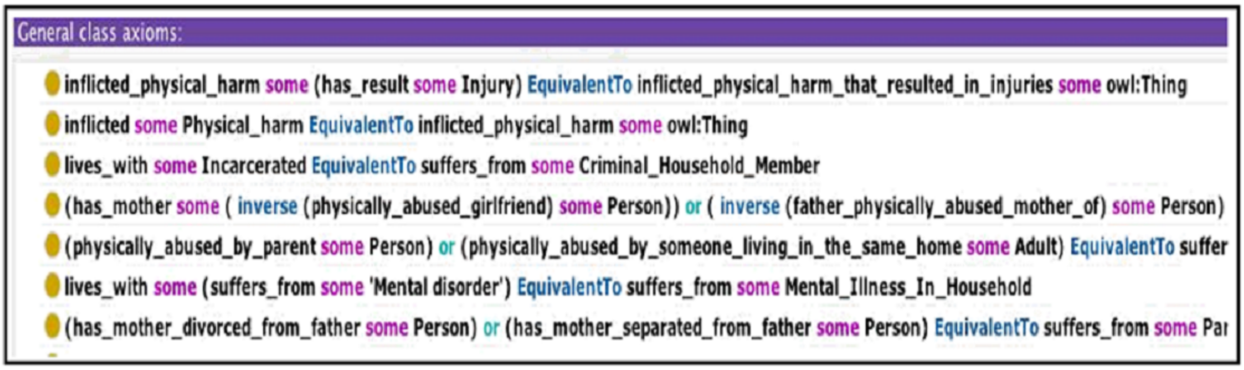
Metrics

Classes	297
Individuals	0
Properties	96
Maximum depth	11
Maximum number of children	29
Average number of children	2
Classes with a single child	81
Classes with more than 25 children	1
Classes with no definition	297

Submissions

Version	Released	Uploaded	Downloads
1 (Parsed, Indexed, Metrics, Annotator)	12/21/2018	12/21/2018	OWL | CSV | RDF/XML

Figure 5. Some general axioms defined in the Adverse Childhood Experiences (ACEs) Ontology. OWL: Web Ontology Language.



```
General class axioms:
inflicted_physical_harm some (has_result some Injury) EquivalentTo inflicted_physical_harm_that_resulted_in_injuries some owl:Thing
inflicted some Physical_harm EquivalentTo inflicted_physical_harm some owl:Thing
lives_with some Incarcerated EquivalentTo suffers_from some Criminal_Household_Member
(has_mother some ( inverse (physically_abused_girlfriend) some Person)) or ( inverse (father_physically_abused_mother_of) some Person)
(physically_abused_by_parent some Person) or (physically_abused_by_someone_living_in_the_same_home some Adult) EquivalentTo suffer
lives_with some (suffers_from some 'Mental disorder') EquivalentTo suffers_from some Mental_Illness_In_Household
(has_mother_divorced_from_father some Person) or (has_mother_separated_from_father some Person) EquivalentTo suffers_from some Par
```

In addition to the actual ontology, we also use semantic rules. In our ontology, some rules are used to enable data access. While the ontology provides a language to express the knowledge, the actual data itself is usually stored in data files or repositories that can use a different lexicon. Rules are employed to map the actual data to the ontology to enable knowledge-based querying and inference. Languages, such as Positional-Slotted, Object-Applicative (PSOA) Rule Markup Language (RuleML) [24], can be used to create those rules. Additionally, in our specific application to ACEs, some concepts require more expressivity than can be supported by OWL [21]. For instance, Felitti et al [1] define that one suffers from *Physical Abuse* if and only if a parent, stepparent, or adult living in her/his home throws something at her/him or pushes, grabs, slaps, or hits the person so hard that marks or injuries are caused. Someone suffered from *Physical Abuse* if, for example, the data contains a subgraph similar (ie, homomorphic) to the graph in Figure 6.

It is not possible to express exactly the existence of such a homomorphism in OWL. Several solutions can be considered. For instance, it is not possible to express that "i_p_h_t_r_i_i_t (inflicted_physical_harm_that_resulted_in_injuries_to) ≅ i_p_h_t_r_i_i (inflicted_physical_harm_that_resulted_in_injuries) o targets" but only that "i_p_h_t_r_i_i o targets ⊆ i_p_h_t_r_i_i_t." As this representation does not allow us to fully express the knowledge and check its consistency, we can use an alternative solution by employing rules external to the ontology, such as "Physically Abused(x) → ∃ y, z has parent(x, y) ∧i_p_h_t_r_i_i(y, z) ∧ targets(z, x)," where x and y would span *Persons* and z would be an *Injury*. The right-hand side of the rule can be a logic formula, as in the example, or a query returning a Boolean value (eg, the query shown in Figure 7).

The third application of rules is to build a recommendation system. In that case, the left-hand side of a rule is a recommendation made to the user. Depending on the use case, the rule and the recommendation can take many different forms. For instance, if the ontology is used during an interview about ACEs with a child patient, the rule could be “If the patient's parents are separated or divorced, ask if they are feeling loved.“ These rules are particularly important in the context of ACEs detection due to the sensitive information (eg, mental illness, handicap, and criminal activity) that needs to be approached carefully.

Another key contribution of this kind of rule is to manage and allocate resources. For example, “If the patient suffers from emotional neglect, schedule an appointment with a child psychologist” (see Figure 8). Initially, these rules reflect a common sense reaction, but as the knowledge grows, new rules reflecting new findings can be added manually or procedurally.

By definition, rules are the result of causal reasoning (ie, if there is a match for the left-hand side, effect the right-hand side) and it is thus natural to connect the rules to causal diagrams. In the current state of this project, the rules are generated from common knowledge and currently established standard procedures and diagnostic processes. We employ causal diagrams [25], built from agreed-upon causality relationships that can be used to generate rules. As the goal of our recommender system is to go beyond what is currently known and practiced, we will use a combination of semantic and statistical inference models to infer the causal paths from the data and its use.

Figure 6. A piece of data representing someone (rectangular node) who was physically injured by a parent (round node). i_p_h_t_r_i_i_t: inflicted_physical_harm_that_resulted_in_injuries_to.

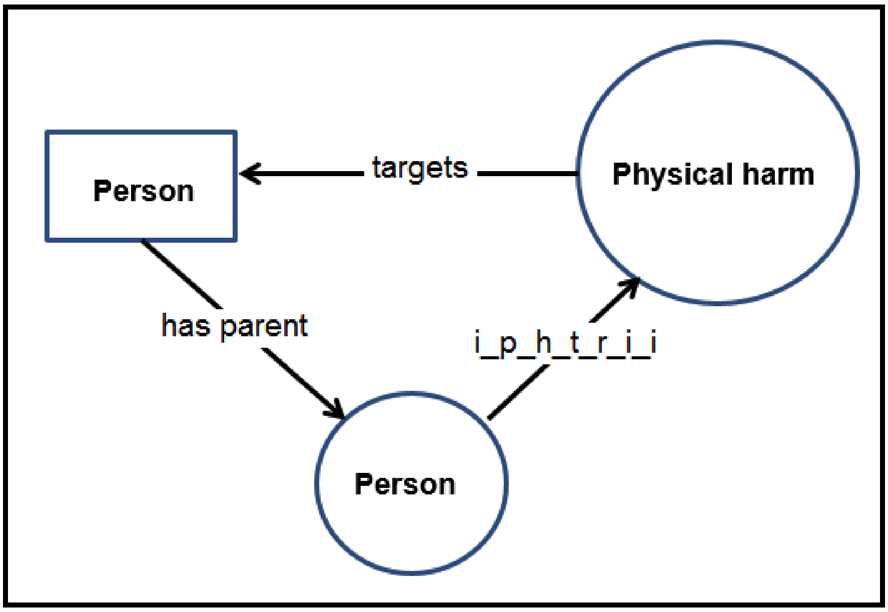
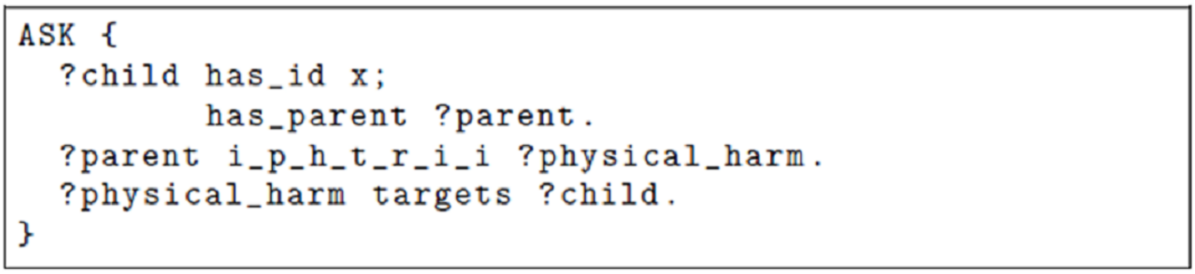


Figure 7. A SPARQL Protocol and Resource Description Framework (RDF) Query Language (SPARQL) query that could be the left-hand side of a rule instantiating the data with Physically Abused(x). i_p_h_t_r_i_i: inflicted_physical_harm_that _resulted_in_injuries.



```
ASK {
  ?child has_id x;
         has_parent ?parent.
  ?parent i_p_h_t_r_i_i ?physical_harm.
  ?physical_harm targets ?child.
}
```

Figure 8. A recommendation rule that schedules an appointment with a psychologist if a patient suffers from emotional abuse (Em_Ab).

Applications

We present three different use-case scenarios for the ACEs Ontology to showcase the role and significance of the ontology to improving ACEs' monitoring, detection, prevention, and management.

Use Case 1: Clinical Setting

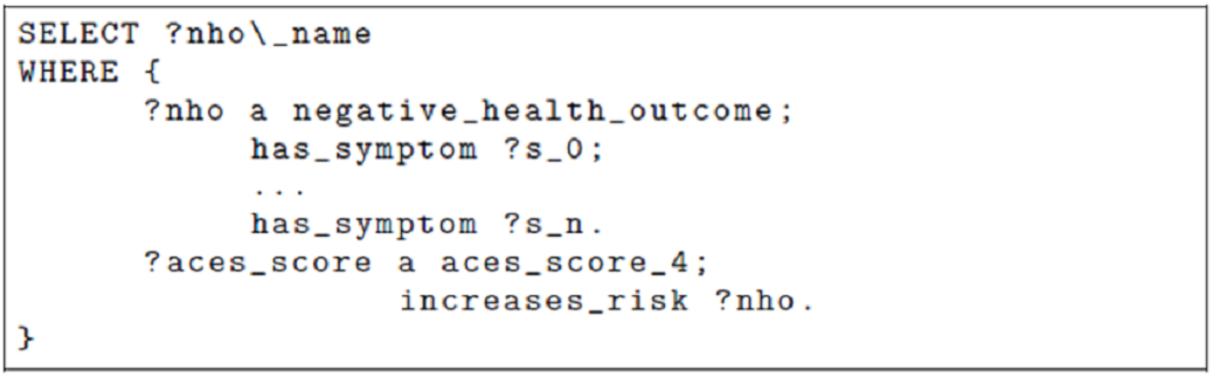
The first use case takes place inside a clinic where a nurse practitioner is interviewing a child patient and her parent. The parent gives personal identification data (ie, name, age, sex, and address) as well as a description of any potential symptoms. This information will be linked to other information that already exists about this patient in the electronic medical record. From the address, using a different data source that gathers data about social determinants of health, the ACEs Ontology helps to infer the patient's social economic status (eg, patient's access to public transportation, the neighborhood's safety and poverty rate, and residential proximity to schools and daycares). The medical practitioner can then use the collected information, as well as the inferred knowledge, to diagnose the illnesses or make a new hypothesis and continue to further investigate the case. Additionally, during the interview, the health practitioner will be able to ask questions relating to ACEs, for example, by slowly introducing household challenges (eg, "Are the parents divorced or separated?" or "Is a household member incarcerated?").

Depending on the answers, using the semantic links in the ontology, the medical practitioner will be able to ask more questions, recommended by the knowledge-based system, to reveal important clues and signs for detection of ACEs. The ontology is used both to allow access to different data sources that share important information and to foster reasoning that will facilitate the knowledge exchange between actors. In this scenario, the goal is mostly to increase the information and knowledge about the patient. Once the data is collected, however, it can be used to formulate a diagnosis. For instance, the ontology can be used to answer the question "Given that the patient has symptoms S_0,..., S_n and an ACEs score of 4, which are the likely negative health outcomes to screen them for?"). An equivalent query in SPARQL Protocol and Resource Description Framework (RDF) Query Language (SPARQL) [26], a standard knowledge-based query language, is shown in Figure 9.

Use Case 2: Public Health Policy Making

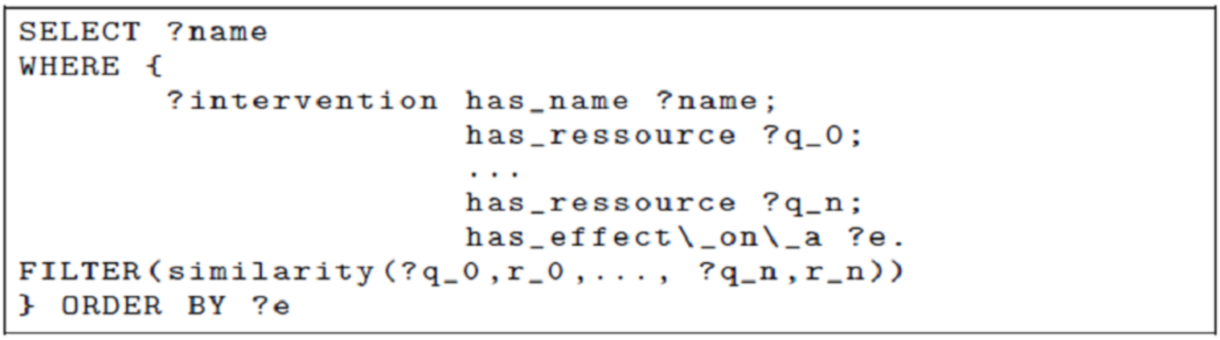
Policy makers and public health organizations can use the SPACES framework for knowledge-based population health surveillance [27] and to identify the causality pathway for the ACEs (eg, physical abuse of spouses and children) and design and implement interventions for remediating them. The semantic framework assists public health planners to compare different communities and programs with each other and finds the intervention (eg, reducing opioid addiction) that is best suited based on each community's needs and priorities. The ontology is used to access, share, and exchange data with other actors, stakeholders, and systems to allow for looking at the problem from various angles and with different granularities to formulate optimal responses. The ontology can then be used to answer questions such as "Given that resources r_0,..., r_n are available, which intervention is most likely to reduce the prevalence of ACE a?" which can be expressed as the SPARQL query shown in Figure 10.

Figure 9. A SPARQL Protocol and Resource Description Framework (RDF) Query Language (SPARQL) query used to discover which negative health outcomes to screen a patient for. nho: negative health outcome; ACEs: Adverse Childhood Experiences.



```
SELECT ?nho\_name
WHERE {
      ?nho a negative_health_outcome;
           has_symptom ?s_0;
           ...
           has_symptom ?s_n.
      ?aces_score a aces_score_4;
                  increases_risk ?nho.
}
```

Figure 10. A SPARQL Protocol and Resource Description Framework (RDF) Query Language (SPARQL) query used to select an appropriate intervention.



```
SELECT ?name
WHERE {
      ?intervention has_name ?name;
                    has_ressource ?q_0;
                    ...
                    has_ressource ?q_n;
                    has_effect\_on\_a ?e.
FILTER(similarity(?q_0,r_0,...,  ?q_n,r_n))
} ORDER BY ?e
```

Use Case 3: Risk Stratification

Another key use of the SPACES framework is in detecting potential problematic areas in order to intervene. As an example, because inflicting abuse is a criminal activity, collecting information on ACEs from the parents can yield less-reliable data than data about neighborhood conditions. It is thus easier to query the data about SDH that are prevalent in the neighborhood in which the patient is living and, thus, which ACEs are the most likely to be a risk. A typical semantic query for such cases can be formulated as “Given that the social determinants of health S_0,..., S_n have been observed in the area, what are the ACEs to screen for?” Similarly, negative health outcomes are easier to track because they often result in hospital visits or medical interventions. The question “Given that the social determinants of health S_0,..., S_n have been observed, and that the negative health outcomes O_0,..., O_m have been frequent in the area, what are the ACEs to screen for?”

Discussion

In this paper, we introduced an ontology for improving the surveillance of ACEs. The goal of the ontology is to provide a uniform structure to represent current and future studies on the causes and effects of, and ways to prevent and mitigate, ACEs. The benefits of using ontologies and semantic technologies have already been shown in several biomedical domains ranging from clinical surveillance [28] to global health surveillance [29]. After describing the architecture of the ontology and discussing the theoretical and engineering aspects involved, we presented multiple applications of the ontology to show how it could contribute to improving the surveillance of ACEs. The ontology was created to be used by major actors of the ACEs community with different applications, from the diagnosis of individuals and the potential negative outcomes that they encounter to the prevention of ACEs in a population and designing interventions and policies. We were able to evaluate the ACEs Ontology automatically using description logic reasoners to ensure its consistency and the satisfiability of its underlying semantic structure. In collaboration with domain experts, we also evaluated the quality of the ontology based on how well the ACEs Ontology aligned with the requirement criteria and standards defined in the project scope statement. Moreover, we assessed the usability of the ontology on the basis of its ability to answer the target queries. Furthermore, along with the further implementation of the SPACES platform, we also intend to design and develop a set of formal evaluation studies for in-depth assessment of the usability of the system and its impact on ACEs’ surveillance and decision making.

The ontology is constantly growing with the advances in knowledge by adding new relevant concepts and relationships as well as new axioms. It is possible for some applications to import only the taxonomy to represent data in a standardized way, while others might use constructs such as transitivity for more complex reasoning. However, the benefits of the ontology are limited by the quality of the data available. Currently, most existing studies deal with adults for whom the negative outcomes are apparent, while more transversal studies of the impact of ACEs on the development of children are lacking [30]. We are currently developing a tool that would use the ontology and semantic inference alongside the statistical inference to assist us in finding patterns in the structured and nonstructured datasets to answer exploratory and explanatory questions.

Acknowledgments

We would like to thank Dr Robert L Davis, Dr Jonathan A McCullers, Dr Sandra R Arnold, and the entire team at the Family Resilience Initiative at Le Bonheur Children's Hospital, Memphis, Tennessee, for their support and insights. This research was supported by the Memphis Research Consortium.

Conflicts of Interest

None declared.

References

1. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 1998 May;14(4):245-258. [Medline: 9635069]
2. Adverse Childhood Experiences: Looking at how ACEs Affect our Lives & Society. Atlanta, GA: Centers for Disease Control and Prevention URL: https://vetoviolence.cdc.gov/apps/phl/images/ACES_Infographic_Accessible.pdf [accessed 2019-03-09] [WebCite Cache ID 76kp7cm7L]
3. Bethell CD, Davis MB, Gombojav N, Stumbo S, Powers K. Issue Brief: A National and Across-State Profile on Adverse Childhood Experiences Among US Children and Possibilities to Heal and Thrive. Baltimore, MD: Johns Hopkins Bloomberg School of Public Health; 2017 Oct. URL: https://www.cahmi.org/wp-content/uploads/2018/05/aces_brief_final.pdf [accessed 2019-03-09] [WebCite Cache ID 76kploSMW]
4. Anda RF, Felitti VJ, Bremner JD, Walker JD, Whitfield C, Perry BD, et al. The enduring effects of abuse and related adverse experiences in childhood. A convergence of evidence from neurobiology and epidemiology. Eur Arch Psychiatry Clin Neurosci 2006 Apr;256(3):174-186 [FREE Full text] [doi: 10.1007/s00406-005-0624-4] [Medline: 16311898]
5. Danese A, McEwen BS. Adverse childhood experiences, allostasis, allostatic load, and age-related disease. Physiol Behav 2012 Apr 12;106(1):29-39. [doi: 10.1016/j.physbeh.2011.08.019] [Medline: 21888923]
6. Garbarino J. ACEs in the criminal justice system. Acad Pediatr 2017;17(7S):S32-S33. [doi: 10.1016/j.acap.2016.09.003] [Medline: 28865658]
7. Ford DE. The community and public well-being model: A new framework and graduate curriculum for addressing adverse childhood experiences. Acad Pediatr 2017;17(7S):S9-S11. [doi: 10.1016/j.acap.2017.04.011] [Medline: 28865666]
8. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. N Engl J Med 2018 Dec 11;379(15):1452-1462. [doi: 10.1056/NEJMra1615014] [Medline: 30304648]
9. World Health Organization. Adverse Childhood Experiences International Questionnaire (ACE-IQ) URL: https://www.who.int/violence_injury_prevention/violence/activities/adverse_childhood_experiences/en/ [accessed 2019-03-09] [WebCite Cache ID 76Ko3Q6Pq]
10. Foundation for Excellence in Mental Health Care. URL: https://www.mentalhealthexcellence.org/projects/adverse-childhood-experiences-prevention-fund/ [accessed 2019-02-20] [WebCite Cache ID 76KoGwXgf]
11. Center for Youth Wellness. URL: https://centerforyouthwellness.org/ [accessed 2019-02-20] [WebCite Cache ID 76KoPWWAH]
12. Tennessee Department of Children's Services. Tennessee ACEs URL: https://www.tn.gov/dcs/program-areas/child-health/aces.html [accessed 2019-02-20] [WebCite Cache ID 76KoUk0X5]
13. Centers for Disease Control and Prevention. Adverse childhood experiences can be prevented URL: https://vetoviolence.cdc.gov/apps/aces/ [accessed 2019-02-20] [WebCite Cache ID 76KoiNgow]
14. ACEs Connection. URL: https://www.acesconnection.com/ [accessed 2019-02-20] [WebCite Cache ID 78Cq8ACbY]
15. Adverse Childhood Experiences. Cumbria Resilience Project URL: https://www.adversechildhoodexperiences.co.uk/about-us [accessed 2019-02-20] [WebCite Cache ID 76KpCus0l]
16. CHEAR Program. URL: https://chearprogram.org/ [accessed 2019-02-20] [WebCite Cache ID 76KpeA39K]
17. BioPortal. Children's Health Exposure Analysis Resource URL: http://bioportal.bioontology.org/ontologies/CHEAR [accessed 2019-02-20] [WebCite Cache ID 76KpvZql1]
18. Brenas JH, Shin EK, Shaban-Nejad A. An ontological framework to improve surveillance of Adverse Childhood Experiences (ACEs). Stud Health Technol Inform 2019;258:31-35. [Medline: 30942708]
19. SNOMED International. URL: https://www.snomed.org/snomed-ct/five-step-briefing [accessed 2018-12-20] [WebCite Cache ID 74oJtTCfG]
20. NCI Thesaurus. 2019. URL: https://ncit.nci.nih.gov/ncitbrowser/ [accessed 2018-12-20] [WebCite Cache ID 74oKOBAx8]
21. MedDRA. URL: https://www.meddra.org [accessed 2018-12-20] [WebCite Cache ID 74oL1zSPB]
22. BioPortal. Adverse Childhood Experiences Ontology URL: https://bioportal.bioontology.org/ontologies/ACESO [accessed 2019-01-14] [WebCite Cache ID 75QM4hXtP]
23. W3C. Web Ontology Language (OWL) URL: https://www.w3.org/OWL/ [accessed 2018-12-20] [WebCite Cache ID 74oLewKx5]
24. Boley H. PSOA RuleML: Integrated object-relational data and rules. In: Faber W, Paschke A, editors. Reasoning Web. Web Logic Rules. Lecture Notes in Computer Science, vol 9203. Cham, Switzerland: Springer International Publishing; 2015:114-150.
25. Okhmatovskaia A, Shaban-Nejad A, Lavigne M, Buckeridge DL. Addressing the challenge of encoding causal epidemiological knowledge in formal ontologies: A practical perspective. Stud Health Technol Inform 2014;205:1125-1129. [Medline: 25160364]
26. W3C. SPARQL query language for RDF URL: https://www.w3.org/TR/rdf-sparql-query/ [accessed 2018-12-20] [WebCite Cache ID 74oLuTUSU]
27. Shaban-Nejad A, Lavigne M, Okhmatovskaia A, Buckeridge DL. PopHR: A knowledge-based platform to support integration, analysis, and visualization of population health data. Ann N Y Acad Sci 2017 Dec;1387(1):44-53. [doi: 10.1111/nyas.13271] [Medline: 27750378]

28. Shaban-Nejad A, Mamiya H, Riazanov A, Forster AJ, Baker CJO, Tamblyn R, et al. From cues to nudge: A knowledge-based framework for surveillance of healthcare-associated infections. J Med Syst 2016 Jan;40(1):23. [doi: 10.1007/s10916-015-0364-6] [Medline: 26537131]
29. Brenas JH, Al-Manir MS, Baker CJO, Shaban-Nejad A. A malaria analytics framework to support evolution and interoperability of global health surveillance systems. IEEE Access 2017;5:21605-21619 [FREE Full text] [doi: 10.1109/ACCESS.2017.2761232]
30. Burke Harris N, Silvério Marques S, Oh D, Bucci M, Cloutier M. Prevent, screen, heal: Collective action to fight the toxic effects of early life adversity. Acad Pediatr 2017;17(7S):S14-S15. [doi: 10.1016/j.acap.2016.11.015] [Medline: 28865647]

Abbreviations

ACE: Adverse Childhood Experience
ACESO: Adverse Childhood Experiences Ontology
CDC: US Centers for Disease Control and Prevention
CHEAR: Children’s Health Exposure Analysis Resource
Em_Ab: emotional abuse
i_p_h_t_r_i_i: inflicted_physical_harm_that _resulted_in_injuries
i_p_h_t_r_i_i_t: inflicted_physical_harm_that_resulted_in_injuries_to
MedDRA: Medical Dictionary for Regulatory Activities
NCBO: National Center for Biomedical Ontology
NCIT: National Cancer Institute Thesaurus
nho: negative health outcome
OWL: Web Ontology Language
PSOA: Positional-Slotted, Object-Applicative
RDF: Resource Description Framework
RuleML: Rule Markup Language
SDH: social determinants of health
SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms
SPACES: Semantic Platform for Adverse Childhood Experiences Surveillance
SPARQL: SPARQL Protocol and RDF Query Language

Edited by J Torous; submitted 25.01.19; peer-reviewed by M Focsa, R Guan; comments to author 14.02.19; revised version received 18.03.19; accepted 02.04.19; published 21.05.19.

Please cite as:
Brenas JH, Shin EK, Shaban-Nejad A
Adverse Childhood Experiences Ontology for Mental Health Surveillance, Research, and Evaluation: Advanced Knowledge Representation and Semantic Web Techniques
JMIR Ment Health 2019;6(5):e13498
URL: http://mental.jmir.org/2019/5/e13498/
doi:10.2196/13498
PMID:31115344

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org