JMIR Mental Health

Journal Impact Factor (JIF) (2023): 4.8
Volume 5 (2018), Issue 3 ISSN 2368-7959 Editor in Chief: John Torous, MD, MBI

Contents

Original Papers

Viewpoint

Reviews

Original Paper

Perceived Attitudes About Substance Use in Anonymous Social Media Posts Near College Campuses: Observational Study

Alexis S Hammond[1], MD, PhD; Michael J Paul[2], PhD; Joseph Hobelmann[1], MD; Animesh R Koratana[3]; Mark Dredze[4], PhD; Margaret S Chisolm[1], MD

[1]Department of Psychiatry and Behavioral Sciences, The Johns Hopkins University, Baltimore, MD, United States

[2]Department of Information Science, University of Colorado, Boulder, Boulder, CO, United States

[3]Department of Computer Science, Stanford University, Stanford, CA, United States

[4]Department of Computer Science, The Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Alexis S Hammond, MD, PhD
Department of Psychiatry and Behavioral Sciences
The Johns Hopkins University
Behavioral Pharmacology Research Unit
5510 Nathan Shock Drive
Baltimore, MD, 21224
United States
Phone: 1 4105500100
Fax: 1 4105500030
Email: ahammon9@jhmi.edu

Abstract

Background: Substance use is a major issue for adolescents and young adults, particularly college students. With the importance of peer influence and the ubiquitous use of social media among these age groups, it is important to assess what is discussed on various social media sites regarding substance use. One particular mobile app (Yik Yak) allowed users to post any message anonymously to nearby persons, often in areas with close proximity to major colleges and universities.

Objective: This study describes the content, including attitude toward substances, of social media discussions that occurred near college campuses and involved substances.

Methods: A total of 493 posts about drugs and alcohol on Yik Yak were reviewed and coded for their content, as well as the poster's attitude toward the substance(s) mentioned.

Results: Alcohol (226/493, 45.8%), marijuana (206/493, 41.8%), and tobacco (67/493, 13%) were the most frequently mentioned substances. Posts about use (442/493) were generally positive toward the substance mentioned (262/442, 59.3%), unless the post was about abstinence from the substance. Additionally, posts that commented on the substance use of others tended to be less positive (18/92, 19.6% positive) compared to posts about one's own use (132/202, 65.3% positive).

Conclusions: This study provides a description of anonymous discussions on or near college campuses about drugs and alcohol, which serves as an example of data that can be examined from social media sites for further research and prevention campaigns.

(JMIR Ment Health 2018;5(3):e52) doi:10.2196/mental.9903

KEYWORDS

Yik Yak; college; social media; substance; drugs; alcohol

Introduction

Substance use is a major public health problem for adolescents and young adults in the United States, and college students have their own particular set of risks for use and barriers to treatment [1,2]. Of an estimated 9 million full-time college students in 2014, approximately 10% reported initial use in the past year of alcohol and 6% reported initial use of illicit drugs (including marijuana, which accounted for the vast majority of this category [3]). Approximately 20% reported using illicit drugs in the past month, almost 40% reported binge drinking, and 13% reported smoking cigarettes. Substance use (particularly alcohol use) has become normative in the college culture, and the influence of

peers is certainly a major factor affecting this stage of development [1].

The use of social media among adults aged 18 to 29 years is now essentially ubiquitous; 90% reported use in 2015 compared to 12% in 2005 [4]. This use has expanded the social network of young adults, and online platforms may have more influence on their substance use than actual in-person interactions [5,6]. Historically, college students have perceived that their peers used substances more than they actually do [7-9]. As such, exposure to substance use via social media may normalize the use for some young adults by presenting it in a positive light, potentially providing a skewed perspective compared to the actual behavior of most college-aged persons. In fact, a study involving the social networking site (SNS) Twitter showed that messages ("tweets") about alcohol use (especially heavy use and binge drinking) are quite common and typically portray a positive attitude about the substance (pro-alcohol tweets were 10 times more frequent than anti-alcohol tweets [10]). Another study examined posts about marijuana on the SNS Instagram, which tends to have younger users (approximately 40% of users are younger than age 24 years), and found that the majority of these posts were also pro-use [11]. Finally, a study examining tweets about menthol cigarettes reported more complex findings, with 48% of tweets being positive and 40% negative, but negative views were common among nonsmokers (91%) and former smokers (71%) [12]. Topics associated with negative sentiment included health and smoking cessation. Together, these findings demonstrate the large variety of substance-related topics discussed on social media sites, which can be a source of data for substance use research, and the complexities of the discussions, which are dependent on the substance mentioned and the perspective of the poster.

Previous studies have shown that substance use rates tend to be higher when collected via self-report methods with some anonymity built in versus direct interviews [13-16], indicating a preference for privacy when discussing this sensitive topic. Accordingly, although some users seem to be open about what they post online, it is possible that certain topics related to substance use may not be as readily discussed by all persons on public sites due to concerns about issues such as privacy or stigma. Thus, it would be interesting to understand whether an added layer of privacy changes the content of posts. The social media mobile app Yik Yak, which was in use from November 2013 until the company ceased operations in May 2017, initially differed from sites such as Facebook because it allowed users to post messages ("yaks") anonymously, without any form of personal identifiers such as usernames, which limited traceability. This was somewhat similar to websites such as Drugs-Forum.com and BlueLight.com, where individuals have partial anonymity (by use of pseudonym usernames) and share information about how to prepare and administer certain drugs, as well as potential effects (positive or negative) to expect. Like Twitter, there was a character limit; Yik Yak had a 200-character limit for each yak. Another difference between Yik Yak and other sites was Yik Yak's restriction of user interaction to a 5-mile radius. Thus, users were only able to communicate with persons that were nearby. The locations popularly served by Yik Yak tended to center around college campuses and, as anyone within a specific area could participate in the network by responding to the post, this effectively targeted a particular audience. Being able to post anonymously to nearby persons not only allowed for free discussion of sensitive topics, such as substance use, but also helped facilitate in-person interactions of people with similar interests. Although these features may have led to the controversial situations and negative press involving Yik Yak due to reports of cyberbullying with subsequent discontinuation of the app [17], research involving SNS or apps with these characteristics provides an opportunity for observation of attitudes and potential practices related to substance use among college students. SNS are very diverse and constantly changing, and it is important to capture information from different sites at varying times to ensure reliability and ease of replication to inform future research.

We previously reported on the types of topics discussed on Yik Yak near college campuses, with a particular focus on general health-related topics [18]. Yik Yak was chosen for study due to the lack of research on it and its features of anonymity and geospatial restriction. We noted that a large share of posts were about sensitive topics such as sex and drugs, which may be related to the anonymity of the platform. Further analysis of substance-related posts showed that most were related to buying substances. In this study, our objective was to further characterize posts that mentioned licit and illicit substances, categorizing them by types of posts, substance(s) mentioned, and poster attitudes toward specific substances while identifying the frequencies of these characteristics. We hypothesized that posts about one's own use of substances would be more likely to result in a categorization implying a positive attitude compared to posts about others' use of substances.

Methods

The dataset was created by downloading messages from Yik Yak. We created a tool that emulated the protocol that mobile devices would use to communicate with Yik Yak servers. This tool allowed us to programmatically retrieve and store yaks in real time into a database for further processing on our end. Additionally, we were able to use the developed tool to change the target location to collect yaks from a variety of locations. We continuously polled the incoming posts, comments, and their respective latitude and longitude for our dataset. The tool allowed us to download messages within a 5-mile radius of a provided latitude and longitude. We used this tool to collect yaks near 120 college campuses in the United States. This set of campuses included the largest universities in the United States, along with additional universities that we included to increase the breadth and diversity of our collection—based on academic rigor, culture, and politics of the location—and also included universities in population-dense cities. For each campus, we queried for yaks within the radius of the campus' geocoordinates, which we obtained from the Google Maps Geocoding API [19]. We downloaded yaks from June 12, 2015 to July 14, 2015, an arbitrary period based on the initial development of our collection tool. We stopped scraping the yaks as our sample size grew to a sizeable amount. The crawler software returned the 100 most recent yaks for a particular location and provided us with 122,179 total yaks.

Table 1. Examples of posts ("yaks") and how they were categorized.

Code	Substance	Paraphrased Yak
Own use, positive	Alcohol	I just want to watch the game, drink some beers, and relax.
Other's use, negative	Tobacco	Cigarette smokers: how selfish to make others breathe in your smoke!
Meeting to use, positive	Marijuana	Who is downtown and wants to smoke a joint with me?
Request for Information, neutral	Other	What's a good price for a gram of cocaine?

Although users could also reply to yaks, we focused on the original yaks in this paper because responses were often quite varied and not always specific to the original post. This study analyzed only the data available from Yik Yak, which maintains the anonymity of the user. The Johns Hopkins University institutional review board approved analysis of publicly available social media messages.

To find yaks relevant to substances, we keyword-filtered original posts matching a large set of substance-related keywords (see Multimedia Appendix 1 for the list of keywords). We manually examined the 12,292 retrieved posts as a first pass and manually removed blatantly irrelevant messages, resulting in a dataset of 2047 substance-related yaks. To characterize the data in greater depth, we reviewed and coded a convenience sample of the first 500 yaks (from a wide variety of universities) of the sample of yaks identified as mentioning substances. Two reviewers (ASH and JGH) read each yak and confirmed whether they were indeed about substances; if they were, the substance(s) was identified and categorized as being alcohol, tobacco, marijuana, and/or other. (The majority of substances fell in the first three categories and due to the small number of other substances mentioned, such as cocaine, Adderall, and LSD, they were grouped together in the "other" category.)

Yaks were then coded for the content based on whether the post was about actual use of a substance or nonuse (typically rhetorical comments about drugs or jokes). Posts specifically about use were then categorized into one of eight types: (1) first-person accounts of use, including effects, (2) comments on use by another person, (3) obtaining substances (through buying or bartering), (4) meeting to use, (5) selling or trading, (6) abstinence from a substance (cessation or cutting down use), (7) laws about use, and (8) questions to obtain information, such as how to use certain substances. Reviewers also coded the displayed attitude (positive, negative, or neutral) of the poster toward the substance(s) mentioned. Reviewers kept this standard by focusing on whether the poster was in support of the substance mentioned specifically, and not the overall emotional affect of the post. If this was ambiguous, or the person was neither in support of nor against the substance, it was rated as neutral. (See Table 1 for paraphrased examples of yaks with codes.) When there was any discordance between the two reviewers at any step in the review and coding process, a third reviewer (MSC) resolved the discordance. The three reviewers established a codebook and discussed what content would be included among each category a priori. Data were analyzed to look at frequencies of particular posts by category and/or substance, and coding agreement among the first two raters was assessed using Cohen kappa coefficient (κ).

Results

Overview

Of the subset of 500 yaks, 493 yaks (98.6%) were confirmed as related to substances on the manual second pass. Although some of the 493 yaks mentioned more than one substance, alcohol (n=226), marijuana (n=206), and tobacco (n=67) were the most frequently mentioned substances. The remaining substances mentioned (n=47) were grouped together as "other" due to low frequency and included "acid," Adderall, methamphetamines, and cocaine. The Cohen kappa score for substance was .98. In all, 53 yaks mentioned two or more substances; those most often mentioned together were alcohol/marijuana (20/53, 38%), alcohol/tobacco (12/53, 23%), alcohol/other substance (11/53, 21%), marijuana/other substance (10/53, 19%), and marijuana/tobacco (7/53, 13%).

Content of Posts

The majority of yaks (442/493, 89.7%; Table 2) were about use of a substance (κ=.56) and, among these, 202 (45.7%) were about the poster's own use, 92 (20.8%) were commenting on someone else's use, 45 (10.2%) involved discussion of meeting up with someone to use, 40 (9.0%) involved buying substances, and 31 (7.0%) asked for information about using. Less common categories of use included the discussion of selling substances (12/442, 2.7%), abstinence from use (9/442, 2.0%), and the legal statuses of substances (10/443, 2.3%). The Cohen kappa score for all categories was .92.

Attitudes of Posts

Overall, posts about substance use were mainly positive toward the substance (262/442, 59.3%; Table 3), with 79 (17.9%) being negative and 101(22.9%) neutral. Tobacco posts tended to be more negative toward the substance, while posts in the "other" category were generally spread out evenly among attitudes. Among the various "use" categories, positive posts included those about meeting up to use (45/45, 100% positive) and about one's own use (132/202, 65.3% positive). Posts about others' use (18/92, 20% positive) and abstinence (1/9, 11% positive) were less positive (Table 4). Posts about the legal status of tobacco were 100% negative (3/3) toward tobacco. These expected attitude trends often persisted when looking at categories of use by substance as well (data not shown), with the notable exception that only 8/32 (25%) comments about other's use of marijuana were negative, with most being positive (10/32, 31%) or neutral (14/32, 43%). The Cohen kappa score for attitudes was .73.

Table 2. Categories of codes, as a function of substance mentioned.

Code	Alcohol (N=226), n (%)	Tobacco (N=62), n (%)	Marijuana (N=206), n (%)	Other (N=47), n (%)
Use	**195 (86.3)**	**62 (92.5)**	**193 (93.7)**	**38 (80.9)**
First-person account of use	104 (53.3)[a]	17 (27.4)[b]	84 (43.5)[c]	21 (55.3)[d]
Comments on others' use	35 (17.9)[a]	31 (50.0)[b]	32 (16.6)[c]	7 (18.4)[d]
Obtaining substance	13 (6.7)[a]	0 (0.0)[b]	22 (11.4)[c]	5 (13.2)[d]
Meeting to use	21 (10.8)[a]	2 (3.2)[b]	24 (12.4)[c]	1 (2.6)[d]
Selling or trading	6 (3.1)[a]	0 (0.0)[b]	6 (3.1)[c]	1 (2.6)[d]
Abstinence	1 (0.1)[a]	7 (11.3)[b]	1 (0.05)[c]	0 (0.0)[d]
Laws about use	1 (0.1)[a]	3 (4.8)[b]	7 (3.6)[c]	0 (0.0)[d]
Requests for information	13 (6.7)[a]	2 (3.2)[b]	17 (8.8)[c]	3 (7.9)[d]

[a]N=195.
[b]N=62.
[c]N=193.
[d]N=38.

Table 3. Attitudes of posters about particular substances.

Attitude toward substance	Alcohol (n=195), n (%)	Tobacco (n=62), n (%)	Marijuana (n=193), n (%)	Other (n=38), n (%)
Positive	116 (59.5)	13 (21.0)	137 (71.0)	18 (47.4)
Negative	30 (15.4)	36 (58.1)	15 (7.8)	9 (23.7)
Neutral	49 (25.1)	13 (21.0)	41 (21.2)	11 (28.9)

Table 4. Attitudes of posters toward substance as a function of the selected category of use.

Attitude	Own use (n=202), n (%)	Others' use (n=92), n (%)	Looking to buy (n=40), n (%)	Requests for information (n=31), n (%)	Abstinence (n=9), n (%)
Positive	132 (65.3)	18 (19.6)	34 (85.0)	16 (51.6)	1 (11.1)
Negative	26 (12.9)	44 (47.8)	0 (0.0)	0 (0.0)	6 (66.7)
Neutral	44 (21.8)	30 (32.6)	6 (15.0)	15 (48.4)	2 (22.2)

Discussion

Our finding that most posts related to substances were positive is consistent with previous studies [10,20-23]. A selection bias exists in this data because these attitudes may not be consistent with the majority of college students, but the findings certainly highlight the type of content to which any college student may be exposed to via social media. A previous study identified prescription drug "abusers" on Twitter and found that persons in their social circles also tended to discuss prescription drug use online [24], providing further evidence of the influence from, and reinforcement of, online content. In fact, viewing and posting about substance use appears to correlate with actual use. Research among Twitter users found that exposure to positive messages about alcohol and marijuana was significantly associated with current heavy episodic drinking and current marijuana use, respectively [25]. Similarly, college students younger than age 21 who posted items on their Facebook profile that were related to intoxication showed higher scores related to problem drinking on the Alcohol Use Disorders Identification Test (AUDIT) scale [26] and were more likely to report having an alcohol-related injury in the past year as opposed to students who did not display references to alcohol [27]. This study was unable to correlate substance use with postings due to the anonymous nature of Yik Yak, and more research is needed to better understand and replicate this phenomenon. As new substances or substance use patterns emerge, social media sites continue to provide an opportunity for health surveillance.

Interestingly, attitudes of posts requesting information about substances, such as how to use them, were almost evenly split between positive and neutral suggesting ambivalence among some posters, which could then be influenced by responses to their posts. We did not analyze replies to original posts, although this would be an interesting avenue for future research to see how often comments agreed or disagreed with the original post or provided helpful or harmful information in response. This could also be an opportunity for intervention in the future: to dissuade young adults from initiating use of a substance or to provide evidence-based health information for this vulnerable population.

Another limitation of the study is, due to the anonymous nature of Yik Yak, we do not know any demographic characteristics of the posters, including age or student status, and have no way to tell if any postings were from automated accounts or "bots" [28]. Posts were shared within a 5-mile radius of a university at the time of posting, so this is a study of areas on and around colleges and universities, but it is not necessarily a study of college students. This information is still important due to the potential for other persons to attempt to sell substances to students, expose them to information about substance use, or meet up with them to use. Additionally, we were limited to a 1-month period of time to collect data, which fell during the summer, a time when some college students are not in town. Future studies may collect more varied information by collecting over a longer duration of time or at repeated points in time.

This study provides a glimpse into the discussion of multiple substances on or near college campuses through an anonymous social media mobile app. Social media sites are constantly changing and evolving, and it is important to collect data across different sites over varying times and durations of time to both capture information and produce reliable results that can be replicated [29]. Yik Yak had unique features, mainly being a location-oriented site with strong anonymity, and it is very likely that similar sites may be developed or those with other novel features. Thus, although Yik Yak is no longer in use and this specific study cannot be replicated, it is still important to report the methodology and findings to inform future studies among current social media sites and novel sites that are certain to arise. The data reported here provide an example of the kind of information that can be examined from publicly available social media posts that may inform health prevention and treatment strategies. For example, this kind of information may prove useful for developing public health campaigns relevant to this population, such as dispelling common myths or advising of the consequences associated with use, possession/distribution, or meeting up with strangers to use.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

MD has received consulting fees from Bloomberg LP, and holds equity in Good Analytics Inc and Sickweather Inc. These organizations did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. MJP serves on the advisory board for Sickweather LLC. The other authors have no conflicts to declare.

Multimedia Appendix 1

List of keywords used in Yik Yak search.

[PDF File (Adobe PDF File), 20KB - mental_v5i3e52_app1.pdf]

References

1. Skidmore CR, Kaufman EA, Crowell SE. Substance use among college students. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):735-753. [doi: 10.1016/j.chc.2016.06.004] [Medline: 27613349]
2. Simons-Morton B, Haynie D, Liu D, Chaurasia A, Li K, Hingson R. The effect of residence, school status, work status, and social influence on the prevalence of alcohol use Among Emerging Adults. J Stud Alcohol Drugs 2016 Jan;77(1):121-132 [FREE Full text] [Medline: 26751362]
3. Lipari R, Jean-Francois B. Day in the life of college students aged 18 to 22: substance use facts. In: The CBHSQ Report. Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality; 2016.
4. Perrin A. Pew Research Center. 2015 Oct 08. Social media usage URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [WebCite Cache ID 6waxSY72Y]
5. Moreno MA, Whitehill JM. Influence of social media on alcohol use in adolescents and young adults. Alcohol Res 2014;36(1):91-100 [FREE Full text] [Medline: 26259003]
6. Boyle SC, LaBrie JW, Froidevaux NM, Witkovic YD. Different digital paths to the keg? How exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students. Addict Behav 2016 Jun;57:21-29. [doi: 10.1016/j.addbeh.2016.01.011] [Medline: 26835604]
7. Kilmer JR, Geisner IM, Gasser ML, Lindgren KP. Normative perceptions of non-medical stimulant use: associations with actual use and hazardous drinking. Addict Behav 2015 Mar;42:51-56 [FREE Full text] [doi: 10.1016/j.addbeh.2014.11.005] [Medline: 25462654]
8. Kilmer JR, Walker DD, Lee CM, Palmer RS, Mallett KA, Fabiano P, et al. Misperceptions of college student marijuana use: implications for prevention. J Stud Alcohol 2006 Mar;67(2):277-281. [Medline: 16562410]
9. Larimer ME, Cronce JM. Identification, prevention, and treatment revisited: individual-focused college drinking prevention strategies 1999-2006. Addict Behav 2007 Nov;32(11):2439-2468. [doi: 10.1016/j.addbeh.2007.05.006] [Medline: 17604915]

10. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. “Hey everyone, I'm drunk.” An evaluation of drinking-related Twitter chatter. J Stud Alcohol Drugs 2015 Jul;76(4):635-643 [FREE Full text] [Medline: 26098041]
11. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-related posts on Instagram. Prev Sci 2016 Aug;17(6):710-720. [doi: 10.1007/s11121-016-0669-9] [Medline: 27262456]
12. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of menthol cigarettes among Twitter users: content and sentiment analysis. J Med Internet Res 2017 Feb 27;19(2):e56 [FREE Full text] [doi: 10.2196/jmir.5694] [Medline: 28242592]
13. Grucza RA, Abbacchi AM, Przybeck TR, Gfroerer JC. Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys. Addiction 2007 Apr;102(4):623-629 [FREE Full text] [doi: 10.1111/j.1360-0443.2007.01745.x] [Medline: 17309538]
14. Schober S, Caces M, Pergamit M, Brandon L. Effect of mode of administration on reporting of drug use in the National Longitudinal Survey. In: Turner CF, Lessler JT, Gfrorerer J, editors. Survey Measurement of Drug Use: Methodological Studies. Rockville, MD: National Institute on Drug Abuse; 1992:177-219.
15. Gfroerer JC, Hughes AL. The feasibility of collecting drug abuse data by telephone. Public Health Rep 1991;106(4):384-393 [FREE Full text] [Medline: 1908589]
16. Aquilino W. Interview mode effects in surveys of drug and alcohol use: a field experiment. Public Opin Q 1994;58(2):A.
17. Safronova V. The New York Times. 2017 May 27. The rise and fall of Yik Yak, the anonymous messaging app URL: https://www.nytimes.com/2017/05/27/style/yik-yak-bullying-mary-washington.html [WebCite Cache ID 6waxxmUl6]
18. Koratana A, Dredze M, Chisolm M, Johnson M, Paul M. Studying anonymous health issues and substance use on college campuses with Yik Yak. 2016 Presented at: Workshops at the Thirtieth AAAI Conference on Artificial Intelligence; Feb 12-13, 2016; Phoenix, AZ.
19. Google Maps Geocoding API. Get started URL: https://developers.google.com/maps/documentation/geocoding/start [accessed 2018-04-27] [WebCite Cache ID 6z01WG3Mt]
20. Thompson L, Rivara FP, Whitehill JM. Prevalence of marijuana-related traffic on Twitter, 2012-2013: a content analysis. Cyberpsychol Behav Soc Netw 2015 Jun;18(6):311-319 [FREE Full text] [doi: 10.1089/cyber.2014.0620] [Medline: 26075917]
21. Lyons AC, Goodwin I, McCreanor T, Griffin C. Social networking and young adults' drinking practices: innovative qualitative methods for health behavior research. Health Psychol 2015 Apr;34(4):293-302. [doi: 10.1037/hea0000168] [Medline: 25822049]
22. Beullens K, Schepers A. Display of alcohol use on Facebook: a content analysis. Cyberpsychol Behav Soc Netw 2013 Jul;16(7):497-503. [doi: 10.1089/cyber.2013.0044] [Medline: 23617225]
23. Allem J, Ramanujam J, Lerman K, Chu K, Boley CT, Unger JB. Identifying sentiment of hookah-related posts on Twitter. JMIR Public Health Surveill 2017 Oct 18;3(4):e74 [FREE Full text] [doi: 10.2196/publichealth.8133] [Medline: 29046267]
24. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
25. Cabrera-Nguyen EP, Cavazos-Rehg P, Krauss M, Bierut LJ, Moreno MA. Young adults' exposure to alcohol- and marijuana-related content on Twitter. J Stud Alcohol Drugs 2016 Mar;77(2):349-353 [FREE Full text] [Medline: 26997194]
26. Schmidt A, Barry KL, Fleming MF. Detection of problem drinkers: the Alcohol Use Disorders Identification Test (AUDIT). South Med J 1995 Jan;88(1):52-59. [Medline: 7817228]
27. Moreno MA, Christakis DA, Egan KG, Brockman LN, Becker T. Associations between displayed alcohol references on Facebook and problem drinking among college students. Arch Pediatr Adolesc Med 2012 Feb;166(2):157-163 [FREE Full text] [doi: 10.1001/archpediatrics.2011.180] [Medline: 21969360]
28. Allem J, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. J Med Internet Res 2016 Aug 09;18(8):e219 [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]
29. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for coding tobacco-related Twitter data: a systematic review. J Med Internet Res 2017 Mar 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
SNS: social networking site

Edited by G Eysenbach; submitted 21.01.18; peer-reviewed by JP Allem, P Chai, K Champion; comments to author 08.04.18; revised version received 01.05.18; accepted 29.05.18; published 02.08.18.

Please cite as:
Hammond AS, Paul MJ, Hobelmann J, Koratana AR, Dredze M, Chisolm MS
Perceived Attitudes About Substance Use in Anonymous Social Media Posts Near College Campuses: Observational Study
JMIR Ment Health 2018;5(3):e52
URL: http://mental.jmir.org/2018/3/e52/
doi:10.2196/mental.9903
PMID:30072359

Original Paper

A Mobile App–Based Intervention for Depression: End-User and Expert Usability Testing Study

Matthew Fuller-Tyszkiewicz[1], PhD; Ben Richardson[2*], PhD; Britt Klein[3*], DPsych (Clinical); Helen Skouteris[4,5*], PhD; Helen Christensen[6*], PhD; David Austin[1*], DPsych (Clinical); David Castle[7*], PhD; Cathrine Mihalopoulos[8*], PhD; Renee O'Donnell[1*], BSc (Hons); Lilani Arulkadacham[5*], PhD; Adrian Shatte[3*], PhD; Anna Ware[1], BSc (Hons)

[1]School of Psychology, Deakin University, Geelong, Australia
[2]Cairnmillar Institute, Melbourne, Australia
[3]School of Science, Engineering & Information Technology, Federation University, Ballarat, Australia
[4]School of Psychology, Deakin University, Melbourne, Australia
[5]School of Psychological Sciences, Monash University, Melbourne, Australia
[6]Black Dog Institute, University of New South Wales, Sydney, Australia
[7]Department of Psychiatry, University of Melbourne, Melbourne, Australia
[8]School of Health and Social Development, Deakin University, Geelong, Australia
[*]these authors contributed equally

Corresponding Author:
Matthew Fuller-Tyszkiewicz, PhD
School of Psychology
Deakin University
1 Gheringhap Street
Geelong, 3220
Australia
Phone: 61 392517344
Email: matthewf@deakin.edu.au

Abstract

Background: Despite the growing number of mental health apps available for smartphones, the perceived usability of these apps from the perspectives of end users or health care experts has rarely been reported. This information is vital, particularly for self-guided mHealth interventions, as perceptions of navigability and quality of content are likely to impact participant engagement and treatment compliance.

Objective: The aim of this study was to conduct a usability evaluation of a personalized, self-guided, app-based intervention for depression.

Methods: Participants were administered the System Usability Scale and open-ended questions as part of a semistructured interview. There were 15 participants equally divided into 3 groups: (1) individuals with clinical depression who were the target audience for the app, (2) mental health professionals, and (3) researchers who specialize in the area of eHealth interventions and/or depression research.

Results: The end-user group rated the app highly, both in quantitative and qualitative assessments. The 2 expert groups highlighted the self-monitoring features and range of established psychological treatment options (such as behavioral activation and cognitive restructuring) but had concerns that the amount and layout of content may be difficult for end users to navigate in a self-directed fashion. The end-user data did not confirm these concerns.

Conclusions: Encouraging participant engagement via self-monitoring and feedback, as well as personalized messaging, may be a viable way to maintain participation in self-guided interventions. Further evaluation is necessary to determine whether levels of engagement with these features enhance treatment effects.

(JMIR Ment Health 2018;5(3):e54) doi:10.2196/mental.9445

KEYWORDS

depression; eHealth; mHealth; young adult

Introduction

Background

Mental illness is both common and costly [1]. People with mental illness have a lower quality of life and poorer work opportunities, and are more likely to attempt suicide [2,3]. Given that onset of depression most often occurs during adolescence to early adulthood [4,5] and that proven treatments for depression become less effective as the severity and duration of the illness increase [6], early detection and treatment of depressive symptoms are seen as key strategies to reduce the prevalence, duration, and burden of depression. Unfortunately, many adolescents and young adults are reluctant to seek help for their depressive symptoms [7], citing barriers such as cost, access to help, lack of anonymity, and perceived stigma of mental illness [8,9]. eHealth technology may help to overcome these barriers to treatment and ultimately assist sufferers in alleviating their depressive symptoms.

eHealth Interventions for Depression

The past decade has seen increasing use of technology to enhance access to, and engagement with, treatments previously established as efficacious when delivered face-to-face [10,11]. This widespread promotion of eHealth solutions to intervention delivery is premised on the notions that eHealth interventions (1) can replicate treatment gains observed in face-to-face treatment; (2) are acceptable to end users; (3) may have advantages in terms of reach, anonymity, and cost; and (4) provide useful features that may enhance treatment experience and outcomes, such as feedback functions to enable charting of one's progress, and sophisticated algorithms for tailoring the intervention experience [12,13].

Accumulated evidence broadly supports these premises. Most adolescents and young adults own a smartphone [14], have accessed a mHealth app [15], and report a stated preference for engaging health services in this manner [16]. Importantly, recent meta-analyses suggest that psychological intervention content delivered via a Web- or mobile app can be as efficacious as face-to-face treatment for depression [10,17,18].

Despite these encouraging findings, it is also clear that dropout rates tend to be higher for eHealth interventions than for face-to-face therapy, especially when eHealth interventions are self-guided [19]. The greater dropout rate in self-guided eHealth interventions may signal that the apps are not sufficiently engaging and/or user-friendly to maintain participant interest over time. Check-ins with a clinician or researcher may enable evaluation of whether the content is being used appropriately, whether it is having the desired effect, and whether there is a need to modify the treatment plan. In instances where clinician contact is not feasible, incorporation of persuasive design principles into app development may enhance end-user experience and outcomes for self-guided apps.

Design Principles to Facilitate Target Behavior

According to the Fogg Behavior Model [20], engagement in a task is dependent on 3 key factors: (1) motivation, (2) sufficient ability for task performance, and (3) triggering to perform the task. This model further stipulates that all the 3 factors are necessary to enable behavior. For instance, motivation alone may be insufficient to bring about change in an individual who does not possess the necessary skills to achieve the desired goal (eg, coping strategies to deal with negative mood). Similarly, triggering performance—whether through an external or internal prompt—may not lead to the desired outcome when the individual is insufficiently motivated to engage in the task at that time.

Within the context of eHealth interventions, incorporation of design features such as engaging, actionable, and easy-to-use content may ensure that ability and motivation levels are less of a barrier to use. Furthermore, personalization of the app through tailoring of content, scheduling of information, and provision of feedback to help participants chart their progress may provide timely prompts that train individuals when to use the content for optimal results. Such features may increase ability and enhance user experience and engagement, and in turn reduce the likelihood of dropout in self-guided treatments [21].

Ultimately, the success of these design choices is determined through usability testing. Jake-Schoffman et al [22] emphasize that usability may be evaluated across a range of dimensions, including how easy the app is to operate, understand, or learn; satisfaction with the app; attractiveness of the layout; and error rates compared with intended usage. Depending on the dimensions one wishes to study, usability may be evaluated using laboratory-based testing [23,24], field testing [25], and/or user feedback [26-28]; however, self-reported usability (hereafter, labeled *perceived usability*) is the most commonly employed usability approach to date for eHealth apps [29].

The Study

This study sought to evaluate end-user experience of a mobile app-based intervention for depression (BlueWatch) developed with feedback functionality as a central feature to enhance usability and engagement. The app was designed to be self-paced and without therapist input and was targeted at young adults (18 to 25 years), given this is a peak period for the onset of depressive symptoms. Although BlueWatch shares many design features with other commonly available depression treatment apps (evidence-based content, survey, mood monitoring features, etc), a key differentiator is that it uses the participants' mood survey data to provide real-time messages to help participants work out when best to engage the treatment content.

A mixed methods design was employed to augment quantitative ratings of perceived usability of BlueWatch with qualitative interviewing to flesh out responses to these questions. These questions probed the ease of use and navigability, aesthetic features of the app, integration and suitability of content provided, and app personalization.

The qualities that make an app attractive for users may differ from features that researchers focus on when developing an intervention or the features that clinicians look for in recommending apps to patients. Hence, the second aim of this study was to compare qualitative responses of the user group with app feedback from clinicians with expertise in the treatment

of depression and researchers who specialize in mental health intervention research (including a focus on eHealth interventions).

Methods

Participants

A total of 5 participants, each from 3 different target groups, were recruited. The majority of participants were female (4 out of 5 for the user and mental health professional groups and 3 out of 5 for the researcher group). The user group had a lower average age (mean 22.4, SD 2.71) than the mental health professional (mean 31.8, SD 6.61) and researcher groups (mean 33.4, SD 5.03).

The target groups included users who had completed use of BlueWatch as part of a randomized controlled trial for individuals with depression (ACTRN12615001093572), mental health clinicians, and mental health researchers. Diagnosis of depression among participants from the user group was ascertained by prescreening with the Patient Health Questionnaire 9 [30], followed by confirmation of diagnosis using the Mini-International Neuropsychiatric Interview [31].

Formative usability trials have demonstrated that a sample of 5 participants can identify 80% of usability issues [32,33], and thus, this study was suitably powered to identify user issues for BlueWatch, both within and across the 3 groups tested. Furthermore, in this study, saturation was reached in the themes derived from qualitative interviews.

Materials

BlueWatch Mobile App Intervention

BlueWatch, a mobile app that comprises short audio activities, journaling exercises, and self-monitoring functions, was designed to improve the well-being and resilience of adults experiencing depressive symptomatology. The app was developed by a multidisciplinary team comprising psychologists, a psychiatrist, and researchers with expertise in eHealth delivery of interventions for depression.

The app was organized into 6 modules based on the principles of cognitive behavioral therapy (CBT): (1) psychoeducation about depression and introduction to CBT, (2) behavioral activation, (3) cognitive restructuring, (4) problem solving, (5) assertiveness skills, and (6) relapse prevention. Each module is based on empirically validated treatment methods and consists of approximately 30 min of content, including instructive text, audio, and various activities to consolidate techniques learned. An example of the content for the values module is provided in Figure 1 to illustrate.

Figure 1. Content for the values module of BlueWatch.

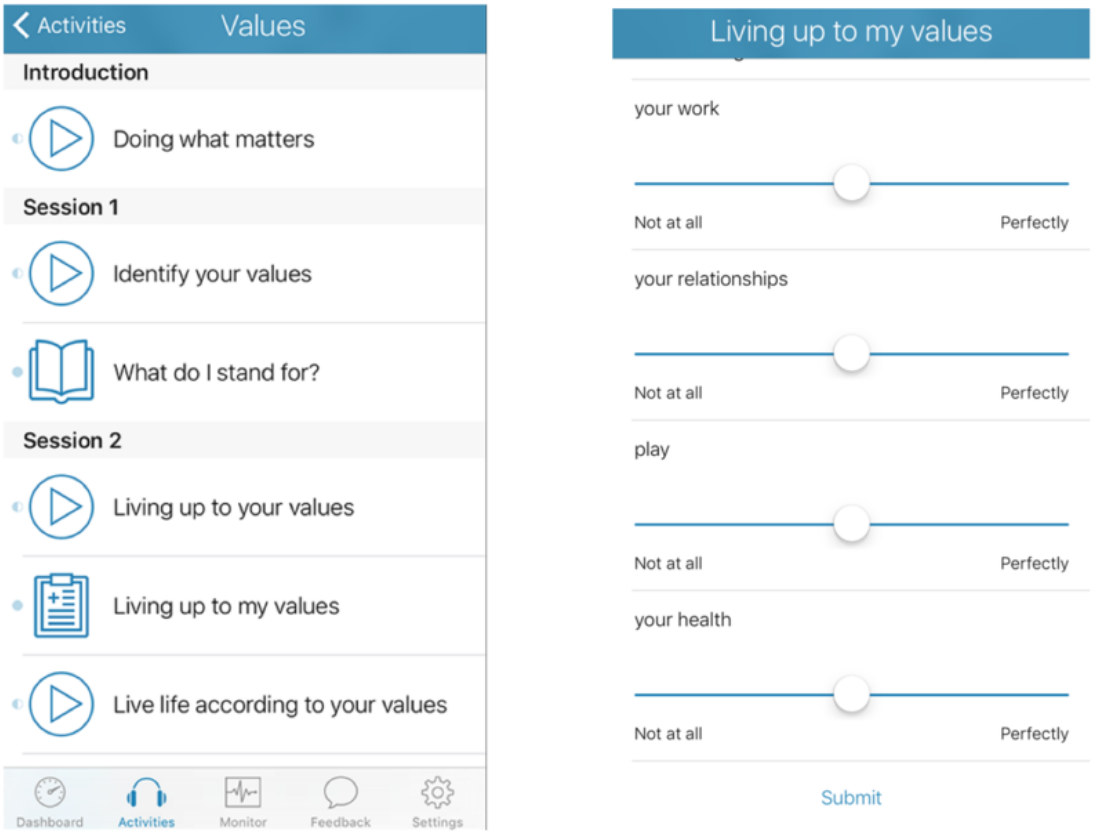
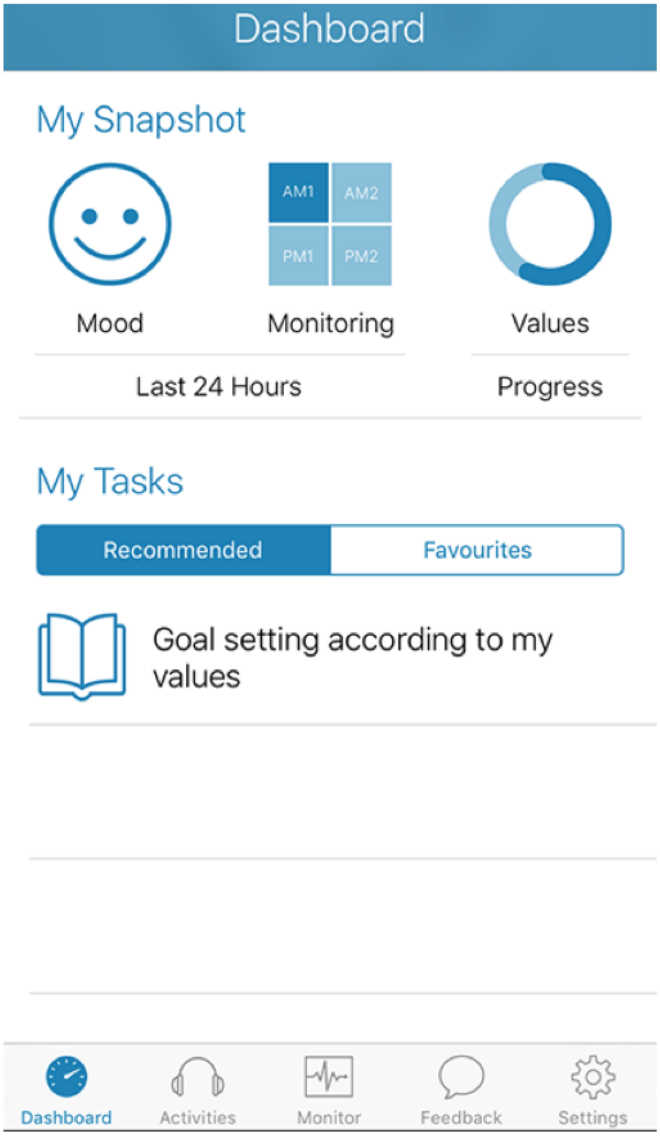


Figure 2. Example of the dashboard used to orient participants to upcoming activities.



Given the self-guided nature of BlueWatch, several design features were added to enhance technologically mediated therapeutic alliance and engagement. The app displays a brief welcome video by default the first time a participant opens the app. This reaffirms the purpose of the intervention, instructs participants how to engage with the app, and points out key features such as push notifications, self-monitoring, and feedback on mood over time. Every time the app is opened thereafter, participants are immediately brought to their dashboard (see Figure 2), which comprises a to-do list, with the option to *favorite* any activities that are liked and may wish to be revisited. As such, the immediate tasks are easy to find and help to prevent participants from losing track of which module they are up to or how to access the relevant content. The list of recommended content is kept brief to prevent overwhelming participants with a list of things yet to be achieved. Furthermore, content is provided in the form of audio files to reduce download size, and a single calming voice is used in each audio file for continuity. Transcripts of audio files are made available (by selecting a button, rather than by default) on screen for each activity to enhance the flexibility of delivery options. Finally, a self-monitoring section is incorporated into the app to allow participants to report their mood states throughout the intervention. These brief surveys take approximately 1 to 2 min to complete and are signaled up to 4 times per day to chart the progress of the participant's mood across the intervention phase (see Figures 3 and 4 for an example of the summary of mood data and questions from the mood survey).

Perceived Usability Measures

Perceived usability was evaluated using the System Usability Scale (SUS; [34]) and a qualitative interview. The SUS is an industry-standard 10-item scale that examines the perceived usability of a technological tool. Responses are measured on a 5-point Likert-type scale with 0 (*strongly disagree*) to 4 (*strongly agree*). Items are summed and then this total is multiplied by 2.5, yielding a SUS composite score between 0 and 100, with higher scores indicating higher perceptions of usability. A SUS score over 68 is considered above average. The SUS has been found as a reliable and valid tool among both

experts and service users when assessing the usability of mobile apps [35].

In addition to the SUS items (which were answered in a phone interview), qualitative questions were included to probe responses to the SUS items as well as to obtain further information about what participants liked and disliked about the app. Structured questions are provided in Multimedia Appendix 1.

Procedure

Recruitment of participants differed depending on the target group they were from. Participants in the user group were recruited after completing the 12-week intervention using BlueWatch. These end users were randomly selected and recruited by invitation via already supplied email addresses, whereas mental health clinicians and researchers were recruited by targeted email invitations that were sent to a number of metropolitan universities and clinical practice centers to recruit relevant experts in the study. Clinicians and researchers were provided with the details to download the BlueWatch app and were instructed to test the app for a period of 7 days before the phone interview.

During the semistructured interview, trained research assistants obtained participant demographics (ie, age and sex) and presented a validated usability scale (the SUS) and open-ended questions to further probe participants' perceptions of the app's usability. The mean interview length was 40 min (SD 9.82), and upon completion, each participant was reimbursed with an AUD $20 shopping voucher. The semistructured interviews were recorded and then transcribed.

Data Analysis

Descriptive statistics were used for quantitative data from the SUS. Results were presented separately for the 3 groups. Both overall SUS scores and item means were reported to provide a more complete picture of perceived usability of BlueWatch. Qualitative data were organized using a coding template analysis approach [36] according to the 6 themes for evaluating the quality of eHealth apps, as proposed in Baumel et al's [37] review of app usability studies: (1) usability, (2) visual design, (3) user engagement, (4) content, (5) therapeutic persuasiveness, and (6) therapeutic alliance. Adoption of a prespecified set of themes rather than deriving themes from interviews permitted prioritization of topics that are widely discussed in the existing literature, thus allowing more direct alignment with this literature. Furthermore, Baumel et al's [37] framework in particular was chosen for several key reasons. First, the framework derives from an extensive review of the existing literature to identify key dimensions of perceived usability. Second, this framework has an emphasis on persuasive design elements and as such aligns with design principles underlying the creation of BlueWatch.

Figure 3. Set up of mood surveys within BlueWatch.

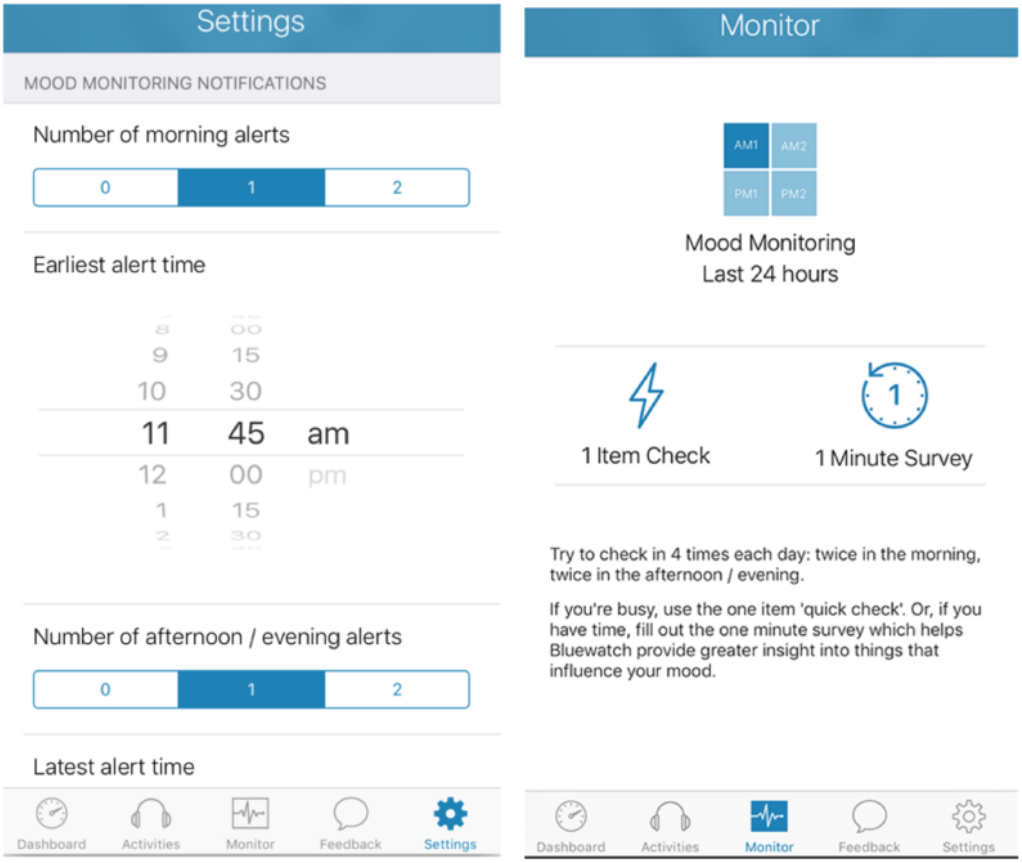
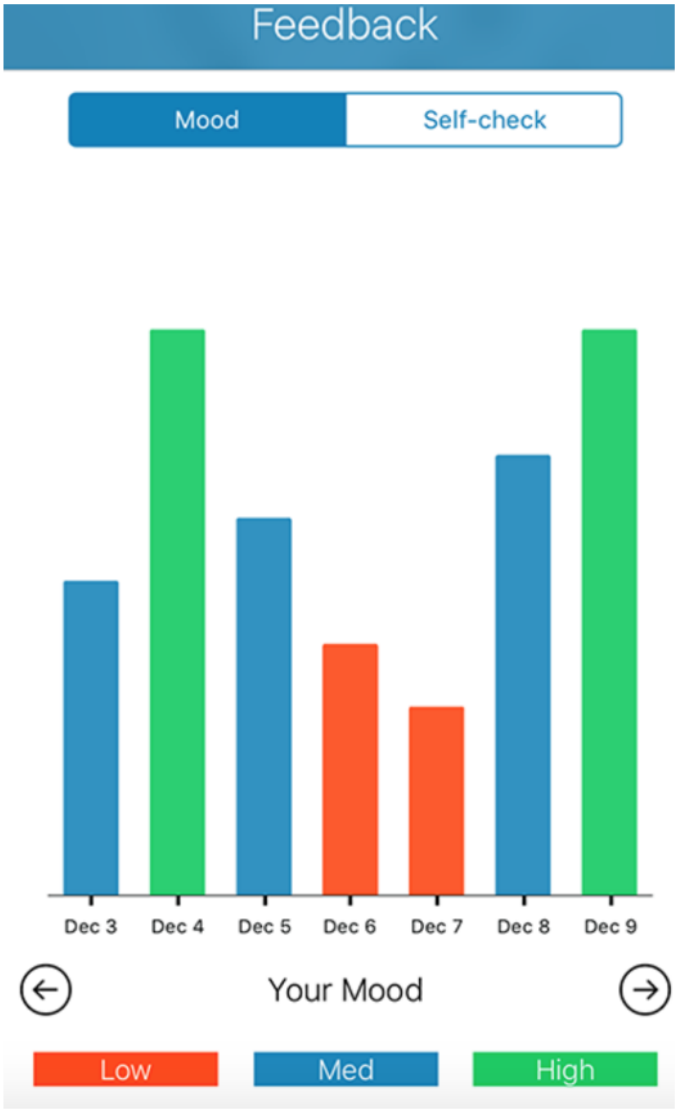


Figure 4. Example of a chart showing mood scores for the past week.



Initially, an additional category for miscellaneous comments was planned for qualitative analysis; however, this seventh category was not needed, given that all comments fit within these 6 initial categories. All coding was done independently by 2 researchers (LA and RO), who then discussed results to ensure consistency and agreement in coded comments. Moreover, 100% consensus was achieved for comment suitability for prespecified themes.

Results

Quantitative Analysis

The quantitative usability data derived from the SUS questionnaire indicated distinctions between researchers and participants. As shown in Table 1, participants tended to report higher usability and positive user experience results than those reported by researchers and clinicians. Although researchers and clinicians tended to have higher scores for items about the difficulty of use and need for support, participants in the end-user group were more likely to strongly endorse items' ease of use and confidence using the app. Overall, the scores suggest that service users rated the app as substantially more usable than clinicians and researchers. However, it is also worth noting that there was more variability in ratings for the researcher and clinician groups, relative to the participant group.

Qualitative Analysis

Content from the semistructured interviews is organized below according to the evaluation categories proposed by Baumel et al [37].

Theme 1: Usability

Service users rated the usability of the app more positively than did the expert groups, although usability comments were generally positive overall. All participants in the user group found the app easy to use and many commented that they found the way in which the activities were presented made sense and "flowed well." This ease of use was further attributed to the welcome message that oriented the users to key features of the app, comparability of the setup of the app with other apps the participants were familiar with, and the *To Do* list as the main page to orient participants to upcoming content:

> *The five slides shown when initially downloading the app was enough to know what to do throughout the app.* [Participant, male]
>
> *If you are familiar with apps then you do not need to learn anything [technically] before using BlueWatch.* [Participant, female]
>
> *As soon as the activities come in, you learn by clicking on them…You do the top one first, and then you go down the list.* [Participant, female]

One participant from the user group also found the option to receive a text transcript of the audio files useful, particularly in situations where playing the audio was not feasible. She emphasized that this option made the app more flexible but cautioned that it took her longer to get through the content when reading and that app users with lower reading ability may not benefit from this option.

Several of the experts expressed that quite some time was required to learn and familiarize themselves with the app:

> *I found the navigation really tricky at the start, if I listen to two CBT sessions and it tells me to go and do an exercise I need to go back into activities. There just seemed like a lot of steps to get my head around at the start.* [Clinician, female]

Although the participants who were interviewed found the language easy to follow, a useful suggestion was made by a researcher to include a glossary of terms within the app in case participants are unfamiliar with or forget the meaning of key terms:

> *Perhaps there could be a separate tab under the activities in which you could include a glossary of key terms that way participants don't need to go back and listen to the audio all over again.* [Researcher, female]

Theme 2: Visual Design

Both the users and experts rated the visual design of the app highly. Users appreciated the well-thought-out appearance of the app and commented that they especially liked its overall simplicity and color palette:

> *It was neutral and straightforward.* [Participant, female]
>
> *I like the blue color, I think it is an attractive app.* [Researcher, female]

Users and experts found the layout and presentation of content "well-organized" and "logical." This was summed up well by a participant from the user group:

> *It was self-explanatory. Going through the content, there weren't that many options…so you couldn't get lost in it.* [Participant, female]

Several interviewees did, however, comment that they would prefer the text to be presented differently:

> *I would suggest increasing the font size—as a 37 year old I found this hard to read.* [Researcher, female]
>
> *When you go to the activities it can be hard to understand as the instructions and your responses are all in the same font. I would change the font or use bold to distinguish.* [Participant, male]

A further suggestion was to use star ratings for mood items instead of the slider scales as currently implemented (Participant, male). This participant further raised the possibility of embedding these star ratings within the activities as well or in addition to the 1-min survey section of BlueWatch.

Theme 3: User Engagement

There was agreement across all participants that the app was highly interactive and personalized, particularly the self-monitoring component that provided users with a graphical representation of their mood states over time and prompts to remind them to use the app (refer to Multimedia Appendix 1).

Table 1. Means and standard deviations for the System Usability Scale across participant, researcher, and clinician groups.

Question[a]	Participants, mean (SD)	Researchers, mean (SD)	Clinicians, mean (SD)
I think I would like to use the app frequently	2.00 (0.7)	2.00 (0.71)	2.20 (0.84)
I found the app to be unnecessarily complex	0.20 (0.44)	1.80 (1.64)	1.20 (1.09)
I thought the app was easy to use	3.40 (0.90)	3.00 (1.22)	3.00 (1.22)
I think that I would need support of a technical person to be able to use the app	0.00 (0.00)	1.40 (1.67)	1.00 (1.22)
I found the various functions in the app were well integrated	2.80 (1.01)	2.20 (0.84)	3.20 (0.45)
I thought there was too much inconsistency in the app	0.60 (0.89)	1.40 (1.14)	0.60 (0.55)
I would imagine that most people would learn to use the app very quickly	3.20 (0.45)	2.80 (1.10)	2.60 (1.14)
I found the app very cumbersome to use	0.40 (0.55)	2.80 (0.84)	2.20 (1.30)
I felt very confident using the app	3.80 (0.45)	3.00 (1.22)	2.40 (0.55)
I needed to learn a lot of things before I could get going with the app	0.40 (0.54)	1.60 (1.50)	1.20 (1.10)
System Usability Scale total score	86.00 (10.84)	60.50 (21.61)	67.00 (15.35)

[a]Responses were scored on a 5-point Likert scale ranging from 0=strongly disagree to 4=strongly agree.

One user noted:

> *Mood tracking was a cool aspect of the app. I treated it like my own personalized diary.* [Participant, female]

There was agreement from all experts that the app achieved a personalized touch as it allowed the user to adjust the app to their unique circumstance. Specifically, there was a function within the app that allowed users to change the time of the day at which alerts were sent (Figure 3), which was found to be very important for this cohort of people who may experience unusual sleeping times:

> *One of the important parts of this is that you can set the alerts to suit your schedule—this is so important for depression in particular because if you have someone who sleeps late they know it is unlikely they will respond to early morning prompts.* [Clinician, female]

Furthermore, experts commented that the graphical feedback was an important function of the app as it allowed users insight into their mood and helped them remember that they may have days on which they feel positive:

> *I really like the idea of being able to see your progress and that you can see how you're going—sometimes people who are depressed think that they are constantly depressed, when they're not, and this lets you see that that is not the case.* [Clinician, female]

However, several experts also noted that a lot of content was delivered via audio (although made available via text transcripts as well) and wondered whether more graphics could be incorporated to help convey the treatment information:

> *If a client learnt visually and understood content through graphs and animations then they would struggle to engage with this app.* [Clinician, male]

Theme 4: Content

There was agreement among all participants that the content presented was rooted in evidence. Indeed, some of the service users commented that they had been exposed to the app components in prior therapeutic exchanges, and this ensured their confidence.

All experts agreed that the content was evidence-based and many had a preference for the behavior activation module, noting its simplicity and clear delivery:

> *I thought the sections on behaviour activation were really well planned out and particularly that is a core treatment of depression regardless of whether you are delving into cognitions or whether you are just looking to change their behaviours. I thought it was a really nice combination of audio plus reading plus the activity.* [Clinician, male]

Clinicians also specifically commented that the variety of evidence-based information presented was impressive as it could cater to a wide range of preferences:

> *I like that you combined CBT with mindfulness—some people just don't like CBT and that is why having mindfulness and also behavioural activation is so important.* [Clinician, female]

Theme 5: Therapeutic Persuasiveness

End-user and expert groups offered differing opinions about the suitability of the amount of content provided by the app. Several experts believed that there were too many activities within the app and that the activities themselves were too long in duration to complete:

> *I think it is too long the activities—it is trying to achieve so much—it is essentially replacing 12 psych session and it seems really ambitious.* [Clinician, female]

In contrast, service users were more accepting of the amount of content and saw this as necessary to provide different strategies to address their depressive symptoms:

> *I didn't find it overwhelming, there wasn't too much information there was the right number.* [Participant, male]

Furthermore, the personalization and self-monitoring aspects of the app seemed to have the desired effect:

> *[The self-monitoring component] made me realise about all the happy small moments that I was having so I notice that I was happier I guess and that my anxieties weren't as big of a deal.* [Participant, female]

Experts agreed on the advantage of providing mood-based feedback but also raised the possibility of incorporating explicit feedback to let participants know once they had completed a module:

> *I wouldn't recommend it for clients who need that feedback immediately—and I find that those with depressive symptoms do often need that to keep up their motivation.* [Clinician, female]

Theme 6: Therapeutic Alliance

However, some users believed the app provided support to an extent but lacked human contact:

> *The app tried to emulate a friend, but it could never actually be that friend.* [Participant, female]

Others felt that the app was an extra resource that they had access to whenever they needed it:

> *I cannot check in with my therapist every day, but I could with the app.* [Participant, female]
>
> *When I felt down, I remembered to use the activities presented in the app to help with the situation.* [Participant, female]

Most experts raised concerns that because of the lack of support offered by a one-on-one therapist, participants could not check their understanding of difficult concepts or points with someone and that this would impact their engagement:

> *I was concerned that there wasn't a section in which a participant could connect with a clinician if they wanted to either solve a problem that emerged from*

using the app or to check understanding of particular concepts. [Clinician, male]

Discussion

Overview

Although eHealth interventions offer a promising way to deliver psychological treatment content, dropout rates for this form of treatment administration may be higher than those for face-to-face therapy [19]. One potential contributing factor for this dropout is difficulty in the use of the app, particularly when the app is designed to be used in a self-guided fashion. In this study, we conducted a usability analysis of an eHealth intervention (BlueWatch) created using persuasive design principles to enhance user experience. Perceived usability data were obtained from end users, clinicians, and mental health researchers to evaluate potential differences in preferences and perceptions of the app.

Principal Findings

There was broad agreement across the 3 groups (end users, health care professionals, and researchers) that the BlueWatch app had an appealing visual layout and organization of content, was engaging for users, and offered evidence-based content that ensured a range of techniques to cater to end users who may prefer different approaches for the treatment of their depressive symptoms. The expert groups (researchers and mental health professionals) particularly liked the wide range of psychological intervention strategies available in the app, feeling that this would allow individuals with different treatment preferences to each find something that might work for them. This eclectic approach is broadly in keeping with recent depression apps, such as MyCompass, that have shown benefit in offering a suite of intervention strategies [38].

A key priority in BlueWatch's design was to leverage self-monitoring of mood as a way to teach participants about their symptoms and to retain engagement in the app. Qualitative feedback from end users suggested that this feature had the desired effect. Participants reported that the mood monitoring surveys and associated graphical feedback were a reason to return to the app and that it increased self-awareness of how their mood fluctuated over time and in relation to use of the intervention content. These findings are consistent with prior research in which use of these app-based interventions promoted increased self-awareness [39,40] and that provision of feedback had positive effects on treatment outcomes [41,42]. Participants also liked that they could adjust the timing and frequency of these mood assessments and felt that the personalized feedback helped to remind them when to use the intervention content.

Given the self-guided nature of BlueWatch, it was not surprising that all groups made mention of the lack of therapist contact. Expert groups raised concern that lack of guidance meant it was unclear whether participants were using the content appropriately. Interviews with the user group did not indicate this to be a problem as they reported that the content was straightforward, they understood how to apply the techniques offered, and found the activities easy to digest. Nevertheless, 1 participant from the user group stated that he would like therapist contact in addition to the app because he saw this contact as an opportunity to discuss actual events in his life rather than relying on exemplars from the app to guide him through daily life circumstances. Contact with the therapist would also enable immediate feedback as he discussed any problems he wished to resolve. To overcome this limitation in BlueWatch's current capabilities, this participant suggested a more interactive element to the journaling task, such that the app might provide feedback on any journal entries to bolster a feeling of connectedness that they might receive in face-to-face therapy.

Limitations

Several characteristics of the present sample may have impacted the results. Participants from the end-user group were invited to participate in the usability interviews if they were aged 18 to 25 years (a peak period for depression), met the diagnostic criteria for depression, and following completion of 12 weeks of use of BlueWatch. Researchers and clinicians were older and likely to have used the app for less time than the end-user group. Such differences may account for discrepancies in ratings and impressions of the app between the end-user and other 2 groups. Furthermore, participants from the end-user group were recruited after completing a 12-week intervention phase. It is possible that these individuals were more motivated and had a more positive experience of the app than those who dropped out of the study.

This study limited its evaluation to perceived usability, and hence, the obtained results are reliant on participants recognizing and conveying any issues they may have in completing app-related tasks. Although this is a common approach to usability testing [29] and the present semistructured interviews sought to evaluate key dimensions of usability, alternative approaches such as lab-based experimental studies [23,24] may offer further insights into whether participants use the app as intended. As a consequence, it remains possible that BlueWatch was positively rated by end users and yet was not used appropriately. Further evaluation across different methods of usability testing is warranted, and may consider including individuals with no prior exposure to the app as well as individuals who have used the app for some time.

Implications

Incorporation of researcher and clinician perspectives demonstrated potential contrast in app experiences relative to end users. The greater perceived ease of use among end users may be due to greater comfort and familiarity with apps in general among younger adults. In the case of researchers, this discrepancy reinforces the need for end-user usability testing to ensure that the proposed features of one's app have the desired effect. In so far as clinicians and researchers are a mechanism for ensuring uptake of apps, the findings of this study suggest that their recommendations might benefit from exposure to end-user feedback as well as their own perceptions of the app.

Conclusions

Usability data from this study broadly supported the use of BlueWatch for treating depression, with particularly positive feedback received from the end-user group. Although

personalization of content through self-pacing, self-monitoring of symptoms, and feedback functions may not substitute for all the benefits of in-person therapist contact, these persuasive design features of BlueWatch appear to enhance engagement with the app's intervention content. This study is part of a larger, ongoing trial of the efficacy of BlueWatch. Results of this evaluation will further illuminate user engagement and usage patterns for those in the intervention group and whether usage rates are associated with symptom reduction postintervention.

Acknowledgments

This work was supported by a grant from Australian Rotary Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questions from semistructured interview of participants.

[PDF File (Adobe PDF File), 22KB - mental_v5i3e54_app1.pdf]

References

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]
2. Keenan-Miller D, Hammen CL, Brennan PA. Health outcomes related to early adolescent depression. J Adolesc Health 2007 Sep;41(3):256-262 [FREE Full text] [doi: 10.1016/j.jadohealth.2007.03.015] [Medline: 17707295]
3. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. Lancet 2012 Mar 17;379(9820):1056-1067. [doi: 10.1016/S0140-6736(11)60871-4]
4. Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
5. McGorry P. Transition to adulthood: the critical period for pre-emptive, disease-modifying care for schizophrenia and related disorders. Schizophr Bull 2011 May;37(3):524-530 [FREE Full text] [doi: 10.1093/schbul/sbr027] [Medline: 21505119]
6. Bukh JD, Bock C, Vinberg M, Kessing LV. The effect of prolonged duration of untreated depression on antidepressant treatment outcome. J Affect Disord 2013 Feb 15;145(1):42-48. [doi: 10.1016/j.jad.2012.07.008] [Medline: 22854096]
7. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007;187(7 Suppl):S35-S39. [Medline: 17908023]
8. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
9. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
10. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
11. Kiropoulos LA, Klein B, Austin DW, Gilson K, Pier C, Mitchell J, et al. Is internet-based CBT for panic disorder and agoraphobia as effective as face-to-face CBT? J Anxiety Disord 2008 Dec;22(8):1273-1284. [doi: 10.1016/j.janxdis.2008.01.008] [Medline: 18289829]
12. Christensen H, Hickie IB. E-mental health: A new era in delivery of mental health services. Med J Aust 2010 Jun 7;192(11 Suppl):S2-S3. [Medline: 20528702]
13. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0]
14. Sensis Social Media Report (1st ed). 2017. URL: https://irp-cdn.multiscreensite.com/535ef142/files/uploaded/Sensis_Social_Media_Report_2017-Chapter-1.pdf [WebCite Cache ID 71W6Vonhx]
15. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

16. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
17. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
18. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
19. 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]
20. Fogg BJ. A behaviour model for persuasive design. 2009 Presented at: Persuasive '09 Proceedings of the 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, California, USA.
21. 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;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
22. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
23. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426. [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
24. Stiles-Shields C, Montague E, Lattie EG, Schueller SM, Kwasny MJ, Mohr DC. Exploring user learnability and learning performance in an app for depression: usability study. JMIR Hum Factors 2017 Aug 11;4(3):e18 [FREE Full text] [doi: 10.2196/humanfactors.7951] [Medline: 28801301]
25. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Experiences and results of applying tools for assessing the quality of a mHealth app named Heartkeeper. J Med Syst 2015 Nov;39(11):142. [doi: 10.1007/s10916-015-0303-6] [Medline: 26345452]
26. Broom MA, Ladley AS, Rhyne EA, Halloran DR. Feasibility and perception of using text messages as an adjunct therapy for low-income, minority mothers with postpartum depression. JMIR Ment Health 2015 Mar;2(1):e4 [FREE Full text] [doi: 10.2196/mental.4074] [Medline: 26543910]
27. Görg N, Priebe K, Deuschel T, Schüller M, Schriner F, Kleindienst N, et al. Computer-assisted in sensu exposure for posttraumatic stress disorder: development and evaluation. JMIR Ment Health 2016 Jun 8;3(2):e27 [FREE Full text] [doi: 10.2196/mental.5697] [Medline: 27277899]
28. Price M, Sawyer T, Harris M, Skalka C. Usability evaluation of a mobile monitoring system to assess symptoms after a traumatic injury: a mixed-methods study. JMIR Ment Health 2016 Jan 11;3(1):e3 [FREE Full text] [doi: 10.2196/mental.5023] [Medline: 26753673]
29. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
31. 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 34-57. [Medline: 9881538]
32. Lewis JR. Sample sizes for usability studies: additional considerations. Hum Factors 1994 Jun;36(2):368-378. [doi: 10.1177/001872089403600215] [Medline: 8070799]
33. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992;34(4):457-468. [doi: 10.1177/001872089203400407]
34. Brooke J. SUS - A quick and dirty usability scale. Usability evaluation in industry 1996;189(194):4-7 [FREE Full text]
35. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. Int J Hum Comput Interact 2013;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
36. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015 Apr 3;12(2):202-222 [FREE Full text] [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]
37. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
38. Proudfoot J, Clarke J, Birch MR, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]

39. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
40. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010 Apr 30;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
41. Hawkins EJ, Lambert MJ, Vermeersch DA, Slade KL, Tuttle KC. The therapeutic effects of providing patient progress information to therapists and patients. Psychother Res 2004;14(3):308-327. [doi: 10.1093/ptr/kph027]
42. Simon W, Lambert MJ, Harris MW, Busath G, Vazquez A. Providing patient progress information and clinical support tools to therapists: effects on patients at risk of treatment failure. Psychother Res 2012;22(6):638-647. [doi: 10.1080/10503307.2012.698918] [Medline: 22755547]

Abbreviations

CBT: cognitive behavioral therapy
SUS: System Usability Scale

Edited by G Eysenbach; submitted 17.11.17; peer-reviewed by S Aggarwal, P Krebs, S March, S Schueller; comments to author 15.03.18; revised version received 10.05.18; accepted 21.06.18; published 23.08.18.

Please cite as:
Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, Castle D, Mihalopoulos C, O'Donnell R, Arulkadacham L, Shatte A, Ware A
A Mobile App–Based Intervention for Depression: End-User and Expert Usability Testing Study
JMIR Ment Health 2018;5(3):e54
URL: http://mental.jmir.org/2018/3/e54/
doi:10.2196/mental.9445
PMID:30139722

Viewpoint

The Effortless Assessment of Risk States (EARS) Tool: An Interpersonal Approach to Mobile Sensing

Monika N Lind[1], MS; Michelle L Byrne[1], PhD; Geordie Wicks[1], BSc; Alec M Smidt[1], MS; Nicholas B Allen[1], PhD

Center for Digital Mental Health, Department of Psychology, University of Oregon, Eugene, OR, United States

Corresponding Author:
Nicholas B Allen, PhD
Center for Digital Mental Health
Department of Psychology
University of Oregon
1227 University Street
Eugene, OR, 97403
United States
Phone: 1 541 346 4075
Email: nallen3@uoregon.edu

Abstract

Background: To predict and prevent mental health crises, we must develop new approaches that can provide a dramatic advance in the effectiveness, timeliness, and scalability of our interventions. However, current methods of predicting mental health crises (eg, clinical monitoring, screening) usually fail on most, if not all, of these criteria. Luckily for us, 77% of Americans carry with them an unprecedented opportunity to detect risk states and provide precise life-saving interventions. Smartphones present an opportunity to empower individuals to leverage the data they generate through their normal phone use to predict and prevent mental health crises.

Objective: To facilitate the collection of high-quality, passive mobile sensing data, we built the Effortless Assessment of Risk States (EARS) tool to enable the generation of predictive machine learning algorithms to solve previously intractable problems and identify risk states before they become crises.

Methods: The EARS tool captures multiple indices of a person's social and affective behavior via their naturalistic use of a smartphone. Although other mobile data collection tools exist, the EARS tool places a unique emphasis on capturing the content as well as the form of social communication on the phone. Signals collected include facial expressions, acoustic vocal quality, natural language use, physical activity, music choice, and geographical location. Critically, the EARS tool collects these data passively, with almost no burden on the user. We programmed the EARS tool in Java for the Android mobile platform. In building the EARS tool, we concentrated on two main considerations: (1) privacy and encryption and (2) phone use impact.

Results: In a pilot study (N=24), participants tolerated the EARS tool well, reporting minimal burden. None of the participants who completed the study reported needing to use the provided battery packs. Current testing on a range of phones indicated that the tool consumed approximately 15% of the battery over a 16-hour period. Installation of the EARS tool caused minimal change in the user interface and user experience. Once installation is completed, the only difference the user notices is the custom keyboard.

Conclusions: The EARS tool offers an innovative approach to passive mobile sensing by emphasizing the centrality of a person's social life to their well-being. We built the EARS tool to power cutting-edge research, with the ultimate goal of leveraging individual big data to empower people and enhance mental health.

(JMIR Ment Health 2018;5(3):e10334) doi:10.2196/10334

KEYWORDS

passive mobile sensing; personal sensing; mobile sensing; mental health; risk assessment; crisis prevention; individual big data; telemedicine; mobile apps; cell phone; depression

Motivation

If health professionals cannot predict and prevent mental health crises, the field faces a crisis of its own. Although we have many evidence-based treatments, greater population-wide access to these treatments has thus far failed to yield significant reductions in the burden of mental health disorders [1-3]. This failure manifests, for example, in the meteoric rise of major depression in the World Health Organization's rankings of conditions responsible for lost years of healthy life [4] and in the recent increase in suicide rates in some 50 World Health Organization member states, including an increase of 28% in the United States from 2000 to 2015 [5,6]. To predict and prevent mental health crises, we must develop new approaches that can provide a dramatic advance in the effectiveness, timeliness, and scalability of our interventions.

Intervening at critical moments in a person's life—that is, during mental health crises, including times of risk for suicide, self-harm, psychotic breakdown, substance use relapse, and interpersonal loss—could dramatically enhance the effectiveness of mental health intervention. Even with the most extreme mental health crises, such as acute suicide risk, we find that the most effective interventions provide barriers to harmful behaviors *at the critical moment of action.* Take, for example, blister packaging on medications commonly used in suicide attempts. By placing a time-consuming, distracting barrier at just the right moment between a person and a suicide attempt, public health policy makers saved lives [7]. If we can improve the timeliness of our interventions, then even low-intensity interventions can have a major impact on improving health and saving lives [8]. Just-in-time adaptive interventions, delivered via mobile health apps and tailored to a person's environment and internal state, headline a host of exciting developments in low-intensity, high-impact interventions [9].

Before we can take full advantage of these approaches, however, we face a critical challenge: *prediction*. It is one thing to recognize the tremendous power of precise timing for interventions, and it is quite another to possess the ability to identify the right time. Furthermore, as we face the challenge of predicting mental health crises, we must commit ourselves to meeting the highest standards of reliability, feasibility, scalability, and affordability. Given the large proportion of individuals experiencing mental health crises who do not receive treatment [10], prediction methods must reach those who are not in traditional mental health care. Current methods of predicting mental health crises (eg, clinical monitoring, screening) usually fail on most, if not all, of these criteria.

Luckily for us, 77% of Americans carry with them an unprecedented opportunity to detect risk states and provide precise life-saving interventions [11]. Smartphones enable the kind of access and insight into an individual's behavior and mood that clinicians dream of. Packed with sensors and integral to many people's lives, smartphones present an opportunity to empower individuals (and their clinicians, if individuals so choose) to leverage the data they generate every day through their normal phone use to predict and prevent mental health crises. This approach to data collection is called passive mobile sensing. Collecting high-quality, passive mobile sensing data will enable the generation of predictive machine learning algorithms to solve previously intractable problems and identify risk states before they become crises. To facilitate this goal, our team built the Effortless Assessment of Risk States (EARS) tool.

Justification

We designed the EARS tool to capture multiple indices of behavior through an individual's normal phone use. We selected these indices based on findings that demonstrate their links to mental health states such as depression and suicidality. These indices include physical activity, geolocation, sleep, phone use duration, music choice, facial expressions, acoustic vocal quality, and natural language use.

Physical activity, geolocation, sleep, phone use duration, and music choice data convey information about how individuals interact with their environments. Physical activity has a rich history of positive outcomes for mental health, including a finding from the Netherlands Mental Health Survey and Incidence Study (N=7076) that showed a strong correlation between more exercise during leisure time and both a lower incidence of mood and anxiety disorders and a faster recovery when those disorders do strike [12]. Geolocation overlaps with physical activity in part, but it can also provide important insight into the quality of a person's daily movement and the environments in which they are spending their time. Saeb and colleagues demonstrated the power of 3 discrete movement quality variables derived from smartphone global positioning system (GPS) data to predict depressive symptom severity [13]. Meanwhile, environmental factors indexed by GPS data, such as living in a city and being exposed to green areas, have consequences for social stress processing and long-term mental health outcomes [14,15]. Sleep provides another powerful signal to help us predict depression and suicidality. One of the most common prodromal features of depression, sleep disturbance (including delayed sleep onset, difficulty staying asleep, and early morning wakening), also relates significantly to suicidality [16,17]. While evidence suggests that phone use duration may affect depression and suicidality via sleep disturbance [18], we believe that phone use duration also deserves study in its own right. For example, Thomée and colleagues found that high phone use predicted higher depressive symptoms, but not sleep disturbance, in women 1 year later [19]. Finally, music choice may provide affective insight, as recent findings suggest that listeners choose music to satisfy emotional needs, especially during periods of negative mood [20].

In contrast with the variables described above, EARS collects other signals that directly reflect both the form and content of an individual's interpersonal engagement. Facial expressions can convey useful information about an individual's mood states, an especially core feature of clinical depression. For example, Girard and colleagues found that participants with elevated depressive symptoms expressed fewer smiles and more signifiers of disgust [21]. Capturing these indicators of mood in a person's facial expressions has become much more efficient and

affordable in recent years with the advent of automated facial analysis [22].

While perhaps less intuitive than facial expression indicators, several aspects of acoustic vocal quality provide robust measures of depression and suicidality. These aspects include speech rate, vocal prosody, vowel space, and other machine learning-derived features [23-26]. Whereas acoustic voice quality ignores the semantic content of communication, natural language processing focuses on it. In addition to strong depression signals in written language in laboratory settings [27], accumulating evidence suggests that social media natural language content, especially expressions of anger and sadness, may identify suicidal individuals [28,29].

We are not the first research team to recognize the potential of passive mobile sensing to capture these behavioral indices. Numerous passive mobile sensing tools capture one or more of the variables described above. Table 1 [30-36] compares the 8 most feature-rich tools (Multimedia Appendix 1 compares the 12 most feature-rich tools [30-40]; for a broader survey of available tools, see the Wikipedia page "Mobile phone based sensing software" [41]).

Passive mobile sensing apps focus on sensor data (eg, GPS, accelerometer) and phone call and text messaging (short message service [SMS]) occurrence (eg, when calls and SMS happen, but not their contents). The EARS tool improves on these apps by adding several indices specifically relevant to interpersonal communication.

Table 1. A comparison of the most feature-rich, research-grade, passive mobile sensing apps[a].

Feature	EARS[b]	AWARE [30]	EmotionSense [31]	Purple Robot [32]	Beiwe [33]	Funf [34]	RADAR-CNS[c] [35]	StudentLife [36]
Geolocation	✔	✔	✔	✔	✔	✔	✔	✔
Accelerometer	✔	✔	✔	✔	✔	✔	✔	✔
Bluetooth colocation		✔	✔[d]	✔	✔	✔	✔	✔
Ambient light	✔	✔	✔	✔			✔	✔
Ambient noise		✔	✔	✔				✔
Charging time	✔							
Screen-on time	✔		✔		✔	✔		
App use	✔	✔	✔[d]	✔[d]		✔	✔	✔
Screen touch events					✔			
SMS[e] frequency	✔	✔	✔	✔	✔	✔	✔	
SMS transcripts	✔			✔				
Call frequency	✔	✔	✔	✔	✔	✔	✔	✔
Browser history						✔		
In-call acoustic voice sample	✔							
All typed text	✔	✔						
Facial expressions	✔							
Music choice	✔							
Barometer		✔						
Wearables	✔						✔	
Video diary[f]	✔							
Audio diary[f]					✔			
Ecological momentary assessment[f]	✔	✔	✔		✔	✔		✔
Count	16	11	10	9	9	9	8	8

[a]EARS, AWARE, EmotionSense, Purple Robot, Beiwe, Funf, and RADAR-CNS are open source.

[b]EARS: Effortless Assessment of Risk States.

[c]RADAR-CNS: Remote Assessment of Disease and Relapse – Central Nervous System.

[d]Availability of the feature is limited to Android versions Kit-Kat 4.4 and below.

[e]SMS: short message service.

[f]Feature is not passive (ie, requires active engagement from the user).

What the Effortless Assessment of Risk States Tool Does

The EARS tool captures multiple indices of a person's social and affective behavior via their naturalistic use of a smartphone. As noted above, these indices include facial expressions, acoustic vocal quality, natural language use, physical activity, music choice, phone use duration, sleep, and geographical location. Critically, the EARS tool collects these data passively, with almost no burden on the user. For example, the EARS tool collects all language typed into the phone. These various data channels are encrypted and uploaded to a secure cloud server, then downloaded and decrypted in our laboratory. Preliminary analyses of these data are underway in our laboratory. Future iterations of the EARS tool will incorporate additional variables and automated analysis on the mobile device itself, which will facilitate both privacy and speed.

The first version of the EARS tool included four key features. A custom keyboard logged every *n* th word typed into the phone across all apps (*n* to be determined by the research team). A patch into the Google Fit application programming interface (Google Inc, Mountain View, CA, USA) collected physical activity data, including walking, running, biking, and car travel. A daily video diary used a persistent notification to prompt users to open the EARS app and record a 2-minute video of themselves talking about their day. While the video diary required active engagement of the user (similar to ecological momentary assessment), it provided a critical bridge in the early EARS tool while our team worked to develop passive means to capture facial expression and acoustic voice data. Each of the above data types (text entry, physical activity, and video diary) were tagged with the final data type of the first EARS tool: geolocation information. Since the Google Fit upload and video diary each occurred only once per day, the keyboard logger provided the richest source of geolocation data. Every time a user entered text into their phone, that text entry triggered a geotag. This approach to gathering geolocation data enabled us to avoid the battery drain of constant GPS data collection.

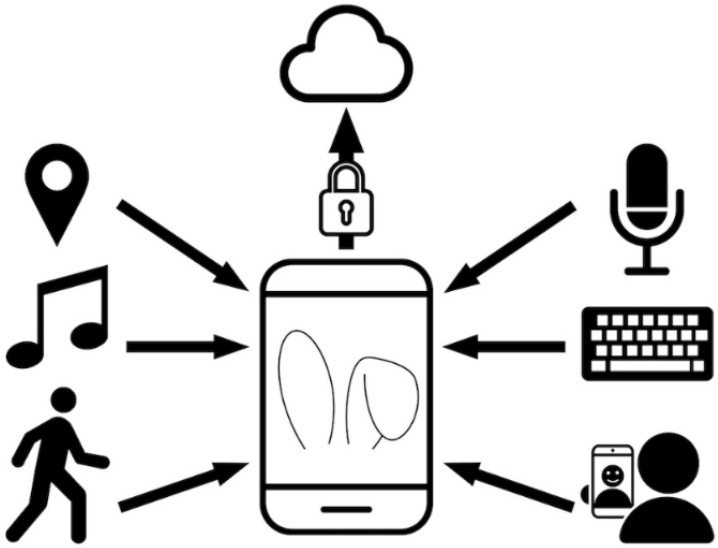
We piloted the first version of the EARS tool in the Effortless Assessment of Stressful Experiences (EASE) study (approved by the University of Oregon Institutional Review Board; protocol number 07212016.019). The EASE study employed an academic stress paradigm to test whether the EARS tool generates data that index stress. We recruited 24 undergraduate students over fall and winter terms of 2016 and 2017 via the Psychology and Linguistics Human Subjects Pool. We acquired informed consent (including descriptions of our double encryption and data storage procedure), then twice collected weeklong sets of passive mobile sensing data. We conducted the baseline assessment 3 to 7 weeks before the students' first final examination (avoiding weeks when they had a midterm or other major project due) and the follow-up assessment during the week prior to the students' last final examination. Self-report questionnaires of perceived stress and mental health symptoms were administered on the last day of each assessment and asked about symptoms over the past week. Participants tolerated the EARS tool well, reporting minimal burden. One individual declined participating in the study due to privacy concerns, and 1 individual declined due to dissatisfaction with the EARS keyboard. One participant dropped out due to privacy concerns, and 1 participant dropped out without explanation. None of the participants who completed the study reported needing to use the provided battery packs.

The current version of the EARS tool includes the four features of the original, plus several enhancements (see Figure 1 [42-51]). The first major enhancement is the addition of a "selfie scraper," which is a pipeline that gathers all photos captured by the device's camera and encrypts and uploads them to our laboratory. During the decryption process, facial recognition software scans the photos and retains only selfies of the participant, discarding all photos that do not include the user's face or that include other faces. This feature enables us to collect facial expression data passively, bringing us a step closer to full passivity. Second, the current version takes another step closer to full passivity with the addition of passive voice collection. The EARS tool records through the device's microphone (but not the earpiece) during phone calls, encrypts these recordings, then uploads them for acoustic voice quality analysis. The third important upgrade of the current version is the constant collection of inertial measurement unit data. These data power fine-grained analysis of physical activity and sleep, over and above what we can glean from Google Fit data. Fourth, paired with the inertial measurement unit data, the ability to sample ambient light via the phone's sensors further enhances the EARS tool's measurement of sleep. The fifth new feature monitors the notification center to capture what music the phone user listens to across various music apps. Sixth, we have added the automatic collection of 4 indices of phone use: SMS frequency, call frequency, screen-on time, and app use time.

The current version of the tool also facilitates integration with wearable technology and adds customization for research teams. To facilitate integration of wearable technology, the EARS tool includes a mechanism to collect raw data from wearable devices (eg, wrist wearables that measure actigraphy and heart rate). This improves efficiency of data collection and reduces the burden on the participant because it cuts out the step of signing in to each participant's individual wearable account to download these data. Capturing raw wearable data may also yield physiological variables that are otherwise obscured by the preprocessing that occurs in standard wearable application programming interfaces. One such variable is respiratory sinus arrhythmia, a measure of parasympathetic nervous system activity that is associated with emotional and mental health states [52].

We hope to share the EARS tool with other research teams, and we recognize that different research questions will call for different passive mobile sensing variables. As such, we are making the EARS tool as modular as possible, so that teams seeking to test hypotheses related to, say, natural language use and selfies do not also have to collect geolocation and app use data. By using Android's flavors functionality, teams can compile a version of EARS that includes the code only for variables of interest, which helps to optimize the tool for battery drain and prevent unnecessary collection of confidential data.

Figure 1. The Effortless Assessment of Risk States (EARS) tool collects multiple indices of behavior and mood.



A person's smartphone use over time generates what we call "individual big data." Individual big data comprise time-intensive, detail-rich data streams that capture the trends and idiosyncrasies of a person's existence. With the EARS tool, we seek to harness these individual big data to power innovation and insight. Our research team aims to detect risk states within participants by determining a person's behavioral set point and analyzing their deviations around that set point. We see this goal as one of many possible applications of the EARS tool.

Effortless Assessment of Risk States Tool Engineering

We programmed the EARS tool in Java for the Android mobile platform (Google Inc). In building the EARS tool, we concentrated on two main considerations: (1) privacy and encryption, and (2) phone use impact. The EARS tool collects a massive amount of personal data, so ensuring that these data remain secure and cannot be used to identify users is of paramount concern. To achieve this, we have implemented a process to deidentify and encrypt the data.

To deidentify the data, the EARS tool uses the Android secure device identification (SSAID) to store and identify the data. The SSAID is accessible only when an Android user gives specific permission and is linked only to the hardware device, not a user or account name. We collect the SSAID and participant name at installation and store the SSAID key on non–cloud-based secure university servers. As a result, in the event of a breach of Amazon Web Services (AWS; Amazon.com, Inc, Seattle, WA, USA), there is no easy way for someone outside of our team to link the data with the name of the user who generated it. Obviously, this basic first step does not protect the actual content of the data. Therefore, we employ state-of-the-art encryption at multiple points in the pipeline.

After the sensors generate the data, the data are immediately encrypted by the EARS tool using 128-bit Advanced Encryption Standard (AES) encryption, a government standard endorsed by the US National Institute of Standards and Technology. On encryption, the unencrypted data are immediately deleted. When transmitting the data to AWS, the EARS tool uses a secure socket layer connection to the server, meaning that all data in transit are encrypted a second time using the industry standard for encrypting data travelling between networks. After transmission to the cloud, the EARS tool then deletes the encrypted data from the phone's memory. On upload to AWS, the data are then also protected by Amazon's server side encryption, which uses 256-bit AES encryption.

By encrypting the user data twice, once on the phone using our own encryption, and a second time on the AWS servers using Amazon's, we can ensure that the data cannot be accessed at any time by anyone outside our team. This means that even Amazon does not have access to the unencrypted data, either by rogue employees or by government demands. We aimed to improve acceptability of the EARS tool for users by taking these steps to protect users' data and allay privacy concerns.

We took another important step toward maximizing acceptability of the EARS tool by prioritizing phone use impact. The EARS tool runs in the background at all times. To minimize the impact on the user's day-to-day experience, we have endeavored to make the EARS tool as lightweight as possible. First, the tool consumes around 30 MB of RAM. Most Android phones have between 2 and 4 GB of RAM, meaning the tool uses between 1% and 2% of memory. Second, the use of phone sensors can have a large impact on the battery life of a phone, as they draw relatively large amounts of power. To combat this, we have moved most cloud uploads to late at night when the phone is usually plugged in, and only when the device is connected to a Wi-Fi network. We also limit GPS readings to once every 5 minutes and, if possible, obtain location data from known Wi-Fi points, rather than connecting to a satellite. Current testing on a range of phones indicated that the tool consumes approximately 15% of the battery over a 16-hour period. Third, installation of the EARS tool causes minimal change in the user interface or user experience. Once installation is completed, the only difference the user will notice is the custom keyboard. With the exception of the optional video diary feature, everything else is collected in the background with no user interaction.

The EARS tool is hosted on GitHub (GitHub, Inc, San Francisco, CA, USA) on the University of Oregon, Center for Digital Mental Health's page (GitHub username: C4DMH). It is licensed under the Apache 2.0 open source license, a permissive free software license that enables the free use, distribution, and modification of the EARS tool [53].

Considerations

We must temper our enthusiasm for the exciting possibilities associated with these new types of data with careful consideration of several challenges. First, EARS is currently implemented only on the Android platform. While Android boasts 2 billion users worldwide, some 700 million people use an iPhone [54]. In the interest of eliminating as many sample selection confounding variables as possible, we have begun adapting the EARS tool for the iOS operating system. Given that most app developers begin on one platform, then port their product over to the other, EARS for iOS should match the functionality of EARS for Android; however, variation in software and sensors will likely result in some differences between the versions. The 4.6 billion people on this planet who do not use an Android or iOS smartphone present a much more stubborn challenge. That most of these people live in low- or middle-income countries or in countries outside of Europe and North America gives us pause, as we consider the historical affluent- and white-centric approach of clinical psychology. The steady increase in smartphone adoption around the world will probably reduce the impact of this limitation, but we remain mindful that the EARS tool carries built-in socioeconomic and cultural limitations alongside its passive mobile sensing features.

Those passive mobile sensing features generate significant ethical concerns as well, principal among which are recruitment and enrollment methods and protecting participants' data. These concerns loom large in mobile health because they reflect the most prominent way in which for-profit app developers exploit smartphone users: app developers bury unsavory content in terms-of-service agreements they know users are unlikely to read, often leveraging these to sell users' data. Some users are aware of these practices and may greet research using the EARS tool with skepticism. However, in our pilot study, only 2 of 28 screened participants declined further participation or dropped out due to privacy concerns. Nevertheless, this issue warrants attention as we scale up our studies and incorporate more diverse populations. The critical task, we believe, is to carefully assess who is empowered by the data and to clearly convey our priorities and precautions to the user. We value the empowerment of the user (eg, to take better control of their mental health) over the empowerment of a commercial or government entity (eg, acquiring user data without any benefit to the user). We oppose practices that empower others at the expense of the user and uphold respect for the user's autonomy by insisting on an opt-in model for all applications of the EARS tool. This means that we reject any protocol that employs autoenrollment in its recruitment approach, such as embedding EARS features into an update of an existing app or requiring members of a specific health plan to participate. To date, we have acquired informed consent in person. As we scale our studies up and out, however, we will need to acquire informed consent remotely. To that end, we are developing a feature in the EARS tool to administer and confirm a participant's informed consent.

Once a participant has opted in, our duty to protect their confidentiality and anonymity grows exponentially. The EARS tool encrypts all data locally on the phone as soon as the user generates them. Those encrypted data arrive in our laboratory via AWS, a US Health Insurance Portability and Accountability Act-compliant commercial cloud service that provides state-of-the-art security in transit. On completion of or withdrawal from a study, a participant's uninstallation of the EARS tool automatically deletes all encrypted EARS data still residing on the phone. As a critical next step to ensure data privacy as we expand this line of research, we aim to conduct processing and analysis locally, on the participant's phone, as soon as possible. In our current protocol, during a study, a participant's encrypted data exist in 3 locations: the phone, the AWS cloud, and our laboratory's secure server. These 3 storage locations increase the risk to the participant. As phones become more powerful and our research generates optimized data analysis algorithms, we aim to limit participants' exposure to risk of privacy breach by executing our protocols within the phones themselves. We will take the first step in that direction soon—recent advances in facial recognition and automatic expression analysis software should enable us to locate the selfie scraper entirely on the phone so that only deidentified output will travel via the cloud to our laboratory [55].

Future Directions

We have described in detail the data collection capabilities of the EARS tool. EARS data, however, are only as useful as the research questions they aim to answer. The next steps for our team include (1) testing the EARS tool with a large, representative sample to establish norms for each behavior, and (2) deploying the EARS tool in the next wave of an ongoing longitudinal study of adolescent girls. We are especially eager to collaborate with adolescent participants because they are digital natives, and we believe their data may be especially revealing. We aim to use the data from these 2 studies to determine which passive mobile sensing variables best predict mental health outcomes. As we discover the predictive power of each variable and the variables' relationships with each other, we will optimize the data pipeline on the backend of the EARS tool to provide meaningful, manageable output for research teams.

Our team built this tool with the goal of predicting and preventing mental health crises such as suicide, but we believe that the EARS tool could serve many uses. The EARS tool could provide rich, observational data in efficacy and effectiveness studies in clinical psychology. It could also index general wellness via fine-grained data on activity, sleep, and mood for public health researchers. Furthermore, if we find that EARS data do predict changes in behavior and mood, medical researchers and health psychologists could use the EARS tool to derive mental and physical health insights. A future version of the EARS tool could provide clinical assessment and actionable feedback for users, including reports that a user could

be encouraged to share with their primary care provider. We imagine exciting applications for social and personality psychologists to test self-report and laboratory findings against unobtrusive, ecologically valid behavioral data. Developmental psychologists, especially those who study adolescence, may also find that the EARS tool provides particularly rich insights, given adolescents' extensive use of mobile computing and their status as digital natives. These applications of the EARS tool depend on rigorous signal processing, exploratory analysis, hypothesis testing, and machine learning methods. The potential is huge, and the work calls for the creative contributions of researchers from myriad areas of study.

Conclusion

Engineers, mobile phone programmers, psychologists, and data scientists have done extraordinary work over the last decade, the sum of which could revolutionize mental health care. We believe that passive mobile sensing could be the catalyst of that revolution. The EARS tool offers an innovative approach to passive mobile sensing by emphasizing the centrality of a person's social life to their well-being. We built the EARS tool to power cutting-edge research, with the ultimate goal of leveraging individual big data to empower people.

Acknowledgments

This work was supported by a grant from the Stress Measurement Network, US National Institute on Aging (R24AG048024). MLB is supported by a US National Institute of Mental Health K01 Award (K01MH111951). The authors would like to thank Benjamin W Nelson and Melissa D Latham for their contributions to this manuscript. The authors also thank Liesel Hess and Justin Swanson for their early contributions to the programming of the EARS tool. Finally, the authors thank the Obama White House 2016 Opportunity Project for facilitating the further development of the EARS tool.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of the 12 most feature-rich, research-grade, passive mobile sensing apps.

[PDF File (Adobe PDF File), 38KB - mental_v5i3e10334_app1.pdf]

References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003 Jun 18;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
2. Mojtabai R, Jorm AF. Trends in psychological distress, depressive episodes and mental health treatment-seeking in the United States: 2001-2012. J Affect Disord 2015 Mar 15;174:556-561. [doi: 10.1016/j.jad.2014.12.039] [Medline: 25556674]
3. Reavley NJ, Jorm AF. Mental health reform: increased resources but limited gains. Med J Aust 2014 Oct 6;201(7):375-376. [doi: 10.5694/mja13.00198] [Medline: 25296048]
4. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
5. World Health Organization. World Health Statistics 2016: Monitoring Health for the Sustainable Development Goals (SDGs) (World Health Statistics Annual). Geneva, Switzerland: World Health Organization; 2016.
6. [No authors listed]. QuickStats: age-adjusted rate* for suicide, by sex - National Vital Statistics System, United States, 1975-2015. MMWR Morb Mortal Wkly Rep 2017 Mar 17;66(10):285 [FREE Full text] [doi: 10.15585/mmwr.mm6610a7] [Medline: 28301452]
7. Yip PSF, Caine E, Yousuf S, Chang S, Wu KC, Chen Y. Means restriction for suicide prevention. Lancet 2012 Jun 23;379(9834):2393-2399. [doi: 10.1016/S0140-6736(12)60521-2] [Medline: 22726520]
8. Johnson EJ, Goldstein D. Medicine. Do defaults save lives? Science 2003 Nov 21;302(5649):1338-1339. [doi: 10.1126/science.1091721] [Medline: 14631022]
9. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
10. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Dec;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
11. Pew Research Center. Mobile fact sheet. Washington, DC: Pew Research Center; 2018. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-02-12] [WebCite Cache ID 6xBeoFSib]

12. Ten Have M, de Graaf R, Monshouwer K. Physical exercise in adults and mental health status findings from the Netherlands mental health survey and incidence study (NEMESIS). J Psychosom Res 2011 Nov;71(5):342-348. [doi: 10.1016/j.jpsychores.2011.04.001] [Medline: 21999978]
13. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
14. Lederbogen F, Kirsch P, Haddad L, Streit F, Tost H, Schuch P, et al. City living and urban upbringing affect neural social stress processing in humans. Nature 2011 Jun 22;474(7352):498-501. [doi: 10.1038/nature10190] [Medline: 21697947]
15. Alcock I, White MP, Wheeler BW, Fleming LE, Depledge MH. Longitudinal effects on mental health of moving to greener and less green urban areas. Environ Sci Technol 2014 Jan 21;48(2):1247-1255. [doi: 10.1021/es403688w] [Medline: 24320055]
16. Lopresti AL, Hood SD, Drummond PD. A review of lifestyle factors that contribute to important pathways associated with major depression: diet, sleep and exercise. J Affect Disord 2013 May 15;148(1):12-27. [doi: 10.1016/j.jad.2013.01.014] [Medline: 23415826]
17. Wojnar M, Ilgen MA, Wojnar J, McCammon RJ, Valenstein M, Brower KJ. Sleep problems and suicidality in the National Comorbidity Survey Replication. J Psychiatr Res 2009 Feb;43(5):526-531 [FREE Full text] [doi: 10.1016/j.jpsychires.2008.07.006] [Medline: 18778837]
18. Lemola S, Perkinson-Gloor N, Brand S, Dewald-Kaufmann JF, Grob A. Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. J Youth Adolesc 2015 Feb;44(2):405-418. [doi: 10.1007/s10964-014-0176-x] [Medline: 25204836]
19. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. BMC Public Health 2011;11:66 [FREE Full text] [doi: 10.1186/1471-2458-11-66] [Medline: 21281471]
20. Randall WM, Rickard NS. Reasons for personal music listening: a mobile experience sampling study of emotional outcomes. Psychol Music 2017;45(4):479-495. [doi: 10.1177/0305735616666939]
21. Girard JM, Cohn JF, Mahoor MH, Mavadati SM, Hammal Z, Rosenwald DP. Nonverbal social withdrawal in depression: evidence from manual and automatic analysis. Image Vis Comput 2014 Oct;32(10):641-647 [FREE Full text] [doi: 10.1016/j.imavis.2013.12.007] [Medline: 25378765]
22. Cohn JF, De la Torre F. Automated face analysis for affective computing. In: Calvo R, D'Mello S S, Gratch J, Kappas A, editors. The Oxford Handbook of Affective Computing. New York, NY: Oxford University Press; 2014:1-58.
23. Cohn JF, Kreuz TS, Matthews I, Yang Y, Nguyen MH, Padilla MT, et al. Detecting depression from facial actions and vocal prosody. 2009 Presented at: Third Annual Conference on Affective Computing and Intelligent Interaction and Workshops; Sep 10-12, 2009; Amsterdam, Netherlands. [doi: 10.1109/ACII.2009.5349358]
24. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
25. Low LA, Maddage NC, Lech M, Sheeber LB, Allen NB. Detection of clinical depression in adolescents' speech during family interactions. IEEE Trans Biomed Eng 2011 Mar;58(3):574-586 [FREE Full text] [doi: 10.1109/TBME.2010.2091640] [Medline: 21075715]
26. Scherer S, Lucas GM, Gratch J, Rizzo AS, Morency LP. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. IEEE Trans Affect Comput 2016 Jan 1;7(1):59-73. [doi: 10.1109/TAFFC.2015.2440264]
27. Rude SS, Gortner EM, Pennebaker JW. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
28. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
29. Coppersmith G, Ngo K, Leary R, Wood A. Exploratory analysis of social media prior to a suicide attempt. 2016 Presented at: 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 16, 2016; San Diego, CA, USA p. 106-117 URL: http://www.aclweb.org/anthology/W16-0311
30. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front ICT 2015 Apr 20;2:6. [doi: 10.3389/fict.2015.00006]
31. Servia-Rodriguez S, Rachuri K, Mascolo C, Rentfrow P, Lathia N, Sandstrom G. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. 2017 Presented at: 26th International World Wide Web Conference; Apr 3-7, 2017; Perth, Australia p. 103-112. [doi: 10.1145/3038912.3052618]
32. CBITs TECH Web Site. Purple Robot URL: https://tech.cbits.northwestern.edu/purple-robot/ [accessed 2018-02-12] [WebCite Cache ID 6xBfQaGeq]
33. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]

34. Aharony N, Gardner A, Saha SK. Funf open sensing. URL: https://github.com/funf-org [accessed 2018-02-12] [WebCite Cache ID 6xBfKoJJJ]
35. Hotopf M, Narayan V. RADAR-CNS. 2016. URL: https://www.radar-cns.org/ [accessed 2018-07-12] [WebCite Cache ID 70qhj5l4v]
36. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Presented at: 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 13-17, 2014; Seattle, WA, USA p. 3-14. [doi: 10.1145/2632048.2632054]
37. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. J Med Internet Res 2017 Mar 16;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
38. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: building a mood sensor from smartphone usage patterns. 2013 Presented at: 11th Annual International Conference on Mobile Systems, Applications, and Services; Jun 25-28, 2013; Taipei, Taiwan p. 389-402. [doi: 10.1145/2462456.2464449]
39. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
40. Apple Inc. ResearchKit. URL: https://developer.apple.com/researchkit/ [accessed 2018-07-12] [WebCite Cache ID 70qho3Yew]
41. Wikipedia. Mobile phone based sensing software URL: https://en.wikipedia.org/wiki/Mobile_phone_based_sensing_software [accessed 2018-07-12] [WebCite Cache ID 70qhVlrkt]
42. Kimmi Studio. Cloud [digital image]. URL: https://thenounproject.com/KimmiStudio/collection/icons/?i=892023 [accessed 2018-02-12] [WebCite Cache ID 6xBgBOWph]
43. Kimmi Studio. Location [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBgHw0G8]
44. Mushu. Lock [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBgSbXEu]
45. Wibowo SA. Microphone [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBgbpczU]
46. i cons. Music [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBgkxeV1]
47. Kimmi Studio. Phone [digital image]. URL: https://thenounproject.com/KimmiStudio/collection/icons/?i=874636 [accessed 2018-02-12] [WebCite Cache ID 6xBfxvJH2]
48. Castillo B. Bunny ears [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBgqcIv0]
49. Hoogendoorn J. Keyboard [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBgxOkL8]
50. Hoffman I. Walking [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBh34cNd]
51. Mikicon. Taking selfie [digital image]. URL: https://thenounproject.com [accessed 2018-02-12] [WebCite Cache ID 6xBh8xYJw]
52. Beauchaine TP. Respiratory sinus arrhythmia: a transdiagnostic biomarker of emotion dysregulation and psychopathology. Curr Opin Psychol 2015 Jun 01;3:43-47 [FREE Full text] [doi: 10.1016/j.copsyc.2015.01.017] [Medline: 25866835]
53. Center for Digital Mental Health. The EARS tool. San Francisco, CA: GitHub Inc; 2018. URL: https://github.com/C4DMH [accessed 2018-02-12] [WebCite Cache ID 6xEYU9Jri]
54. Reisinger D. Here's how many iPhones are currently being used worldwide. 2017 Mar 6. URL: http://fortune.com/2017/03/06/apple-iphone-use-worldwide/ [accessed 2018-02-13] [WebCite Cache ID 6xBhCqOVw]
55. Bader D. The Galaxy S8 has face recognition and iris scanning, and you have to choose one. 2017 Apr 7. URL: https://www.androidcentral.com/galaxy-s8-face-recognition-iris-scanning [accessed 2018-02-13] [WebCite Cache ID 6xBhGxDbD]

Abbreviations

AES: Advanced Encryption Standard
AWS: Amazon Web Services
EARS: Effortless Assessment of Risk States
EASE: Effortless Assessment of Stressful Experiences
GPS: global positioning system
RADAR-CNS: Remote Assessment of Disease and Relapse – Central Nervous System
SMS: short message service
SSAID: secure device identification

Edited by G Eysenbach; submitted 07.03.18; peer-reviewed by K Goniewicz, A Roundtree; comments to author 19.04.18; revised version received 12.07.18; accepted 30.07.18; published 28.08.18.

Please cite as:
Lind MN, Byrne ML, Wicks G, Smidt AM, Allen NB
The Effortless Assessment of Risk States (EARS) Tool: An Interpersonal Approach to Mobile Sensing
JMIR Ment Health 2018;5(3):e10334
URL: http://mental.jmir.org/2018/3/e10334/
doi:10.2196/10334
PMID:30154072

Original Paper

Measurement of Symptom Change Following Web-Based Psychotherapy: Statistical Characteristics and Analytical Methods for Measuring and Interpreting Change

Eyal Karin[1], MApStat; Blake F Dear[1,2], PhD; Gillian Z Heller[3], PhD; Milena Gandy[1], PhD; Nickolai Titov[1,2], PhD

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

[2]Mindspot Clinic, Macquarie University, Sydney, Australia

[3]Department of Statistics, Faculty of Science and Engineering, Macquarie University, Sydney, Australia

Corresponding Author:
Eyal Karin, MApStat
eCentreClinic
Department of Psychology
Macquarie University
Building C3A
First Walk Macquarie University NSW
Sydney, 2109
Australia
Phone: 61 298508657
Email: eyal.karin@mq.edu.au

Abstract

Background: Accurate measurement of treatment-related change is a key part of psychotherapy research and the investigation of treatment efficacy. For this reason, the ability to measure change with accurate and valid methods is critical for psychotherapy.

Objective: The aims of this study were to (1) explore the underlying characteristics of depressive symptom change, measured with the nine-item Patient Health Questionnaire (PHQ-9), following psychotherapy, and (2) compare the suitability of different ways to measure and interpret symptom change. A treatment sample of Web-based psychotherapy participants (n=1098) and a waitlist sample (n=96) were used to (1) explore the statistical characteristics of depressive symptom change, and (2) compare the suitability of two common types of change functions: linear and proportional change.

Methods: These objectives were explored using hypotheses that tested (1) the relationship between baseline symptoms and the rate of change, (2) the shape of symptom score distribution following treatment, and (3) measurement error associated with linear and proportional measurement models.

Results: Findings demonstrated that (1) individuals with severe depressive baseline symptoms had greater reductions in symptom scores than individuals with mild baseline symptoms (11.4 vs 3.7); however, as a percentage measurement, change remained similar across individuals with mild, moderate, or severe baseline symptoms (50%-55%); (2) positive skewness was observed in PHQ-9 score distributions following treatment; and (3) models that measured symptom change as a proportional function resulted in greater model fit and reduced measurement error (<30%).

Conclusions: This study suggests that symptom scales, sharing an implicit feature of score bounding, are associated with a proportional function of change. Selecting statistics that overlook this proportional change (eg, Cohen *d*) is problematic and leads to (1) artificially increased estimates of change with higher baseline symptoms, (2) increased measurement error, and (3) confounded estimates of treatment efficacy and clinical change. Implications, limitations, and idiosyncrasies from these results are discussed.

(JMIR Ment Health 2018;5(3):e10200) doi:10.2196/10200

KEYWORDS

clinical measurement; treatment evaluation; symptom change; symptom scales; psychotherapeutic change

Introduction

Accurate measurement of treatment-related change is a key part of psychotherapy research [1-3] and the investigation of treatment efficacy [4-6]. For example, measurable change in symptoms of anxiety and depression is often used as the primary means to research and test the safety of emerging treatments [7]. Reporting symptom change in anxiety and depression has been shown to describe the clinical trajectory of participants in treatment [8], illustrate the cost-effectiveness of treatment [9], and compare treatments [10]. For this reason, the ability to measure change with accurate and valid methods is critical for psychotherapy [6,11].

Several statistical and clinical methods are employed to increase the validity and accuracy of change measurement in psychotherapy. The most common methodology in psychotherapy research is the combined use of standardized scales, such as standardized symptom scales of anxiety [12] or depression [1,13], and the use of statistical analyses, such as Cohen *d* effect sizes, that measure and interpret the rate of change in treatment [4-6]. Many types of standardized scales are available for measuring and interpreting change in treatment (eg, clinical interviews, measurement of behavior or quality of life [14]), and that change can be statistically estimated through various statistical methods [15]. However, from the wide range of possible methods for measuring treatment outcomes [16], the use of standardized scales, primarily symptom scales, in combination with effect sizes, primarily Cohen *d*, are the most influential. For example, symptom scales and effect sizes are used to evaluate treatment-related change and treatment efficacy within psychotherapy trials [17-19], epidemiological studies [20,21], meta-analytic studies of various treatments [22], and are even mandated within clinical guidelines for reporting in clinical trials, such as Consolidated Standards of Reporting Trials (CONSORT) [19], Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) [23], Strengthening the Reporting of Observational studies in Epidemiology (STROBE) [24], and others [11].

Notwithstanding the common use of both symptom scales and effect sizes for measuring psychotherapeutic-related change, little research is currently available to verify or refute the use of different statistical methods for measuring and interpreting symptom change [25,26]. For example, the use of effect sizes, such as Cohen *d*, is based on statistical assumptions that change is linear. In technical terms, by employing effect sizes, researchers assume that the symptom change that follows treatment is average, constant, and representative of the average change experienced by any participating individual [18,27]. Put another way, if an average individual with moderate depressive symptoms prior to treatment, such as a score between 10 and 15 on the nine-item Patient Health Questionnaire (PHQ-9), would improve by 5 points on a symptom scale, an individual with severe baseline symptoms (eg, PHQ-9 score of 20-27) would be expected to demonstrate the same rate of improvement (eg, 5 points). Similarly, under the linear assumption, a group of participants with different baseline symptoms (eg, mild, moderate, or severe baseline symptoms) undertaking the same therapy would be expected to have similar effect sizes between groups (eg, 1.0). However, in contrast to the common use of statistics that assume change is linear, there are two lines of research to suggest that real-world symptom change may occur as a proportional function from baseline. First, psychological treatment studies often describe an increased rate of clinical change within samples of increased baseline symptom severity [20,28]. Second, common symptom scales, such as the PHQ-9 [29], the Generalized Anxiety Disorder seven-item scale (GAD-7) [30], and prominent others (eg, Kessler Psychological Distress scale) [31], often demonstrate an implicit design feature of score bounding at minimal symptoms. This bounding within symptom scales should theoretically imply that, under effective treatment, all individuals would reduce their symptoms down to the same endpoint of minimal levels [1,9] and that the rate of change would systematically depend on an individual's symptoms at baseline [32,33].

From a statistical point of view, identifying the characteristics of symptom change, and employing a suitable statistical analysis that captures the underlying function of change, can fundamentally impact both the measurement and interpretation of clinical outcomes [15,34,35]. For example, under circumstances in which change is proportional in nature, the selection of a proportional statistical analysis can greatly increase the accuracy and validity of estimating longitudinal clinical change [34,35]; the detection of moderators of symptom change [36]; the classification of subgroups, such as remitters or nonresponders [37]; as well as the ability to research other objectives [38]. For this reason, the function of symptom change must be researched and more clearly understood. Such research could verify, refute, and draw out the implication for using well-established statistical methods (eg, effect sizes, linear statistics) and emerging alternatives (eg, percentage improvement, generalized linear statistics) for measuring and interpreting change in treatment. In addition, researching the function and characteristics of symptom change has the potential to inform researchers and the broader community about the type of change individuals in treatment are likely to experience.

This Study

This study aims to (1) explore the fundamental statistical characteristics of treatment-related depressive symptom change and (2) compare the implications from measuring and interpreting clinical change through effect sizes, such as Cohen *d*, against emerging alternatives, such as percentage improvement (proportional, generalized longitudinal linear statistics) [25,26].

This study employed a large sample of individuals (N=1098) who underwent Web-based psychotherapy (Internet-delivered cognitive behavioral therapy [ICBT]) [39] for symptoms of depression (PHQ-9 [29]). Although Web-based psychotherapy represents a distinct type of psychotherapy, the use of Web-based treatments, which standardizes treatment materials and participant engagement through automatization, can be seen as an opportunity for researching symptom change with high internal validity and minimum methodological interference.

The statistical characteristics of symptom change were explored with three steps. Initially, the relationship between baseline symptoms and the rate of change was explored. In line with

previous clinical studies that suggest that more severely symptomatic participants demonstrate increased effect sizes [20,32], it was hypothesized that individuals with increased symptoms at baseline would also demonstrate increased rates of symptom change (hypothesis 1). Second, the shape of symptom score distribution before and following treatment were explored. In line with the suggestion that symptoms scores are bounded at minimal symptoms [29,30], the distributions of pretreatment and posttreatment depression symptom levels were hypothesized to show evidence of positive skewness and kurtosis at both pretreatment and posttreatment (hypothesis 2). Third, the measurement error associated with linear and proportional measurement models was compared. In line with the characterization of symptom change as proportional, it was hypothesized that those statistical methods that measure symptom change as a proportional function would be associated with reduced measurement error and indicate greater statistical fit to real symptom data in treatment (hypothesis 3). Finally, an additional effort was taken to explore the patterns of depressive symptom change within a control group (n=96). This addition was designed to explore the pattern of symptom change that is not specific to treatment.

Methods

The Sample

This study combined clinical data from three published randomized controlled trials, all of which evaluated ICBT for symptoms of depression and anxiety [39,40]. These interventions were almost identical in structure and therapeutic content. All trials were delivered using the same evidence-based online treatment approach [7] and were conducted within the same research clinic, the eCentreClinic [41]. A precautionary test, aiming to compare the symptom reduction rates between the individual trials, demonstrated similarities across all three interventions. Specifically, a generalized estimated equation (GEE) model [35], testing the longitudinal symptom change of each trial, resulted in slight differences in the estimates of symptom change across trials (PHQ-9 range 5.23-6.29 points); differences were not statistically significant (group × time: Wald $\chi^2_{2,2368}$=5.0, P=.08).

Together, these trials represent a large random intake of self-selecting adults into treatment over a period of 2 years with a total of 1262 adult participants, of whom 1098 (87.01%) were successfully assessed at both pretreatment and posttreatment time points. Additional information about recruitment, advertising, treatment materials, and additional treatment procedures can be found within additional eCentreClinic publications [7,41].

To be included in these trials, participants were selected on the basis of (1) demonstrating at least mild symptoms of depression or anxiety (a minimum score ≥5 on either the PHQ-9 or the GAD-7), (2) older than 18 years and younger than 65 years, (3) being an Australian resident, and (4) having Internet access for the period of the trial. In addition, applicants who reported a score of 3 (considered severe) on item 9 of the PHQ-9 measuring suicidal risk, were referred to another service.

Additional demographic and symptom characteristics are shown in Table 1 for both the treatment and waitlist control conditions.

Table 1. Sample demographics (N=1194).

Demographics	Collated treatment sample (n=1098)	Control sample (n=96)
Gender (male), n (%)	330 (30.1)	51 (53.1)
Age (years), mean (SD)	52.8 (14.2)	56.3 (13.0)
Using medication during the course, n (%)	351 (31.9)	51 (53.1)
Married, n (%)	713 (64.9)	45 (46.9)
Employed, n (%)	636 (57.9)	49 (51.0)
Education, n (%)		
High school	176 (16.0)	39 (40.6)
Vocational education	307 (27.9)	24 (25.0)
Degree	615 (56.0)	37 (38.5)
PHQ-9[a], mean (SD)		
Before treatment	11.73 (4.83)	10.95 (4.73)
following treatment)	5.60 (4.58)	11.00 (5.04)
GAD-7[b], mean (SD)		
Before treatment	10.91 (4.53)	9.5 (4.53)
Following treatment	5.47 (4.35)	8.83 (4.67)

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

[b]GAD-7: seven-item Generalized Anxiety Disorder scale.

Symptom Measure

The PHQ-9 was employed as the primary outcome variable, measuring the presence and severity of depressive symptoms [29]. The PHQ-9 is widely used in clinical trials [7,16], comprising nine items, with high internal consistency and high sensitivity to the presence and change of clinical depression diagnoses [29]. Scores on the PHQ-9 correspond to the cumulative experience of common depressive symptoms over the preceding 2-week period. Cumulative scores range from 0 to 27 and scores are clinically interpreted as falling within five categories: (1) no depression symptoms (total score: 0-4), (2) mild depression symptoms (total score: 5-9), (3) moderate depression symptoms (total score: 10-14), (4) moderately severe depression symptoms (total score: 15-19), and (5) very severe depression symptoms (total scores: 20-27). Symptom scores were modified with a small constant added (0.001) to ensure that plausible values of zero symptoms at posttreatment were represented in the model when statistically modeling proportional functions, such as logarithmic link functions.

Analytical Plan

The function of symptom change was explored with three separate steps, corresponding to the three hypotheses.

The first hypothesis that individuals with increased symptoms at baseline would also demonstrate increased rates of symptom change was tested by examining the relationship between baseline symptoms and the rate of symptom change. Symptom change was examined within the five subgroups of individuals of different baseline PHQ-9 score bands (eg, minimal to no symptoms to very severe depression symptoms). Within each subgroup, the rate of change was approximated with GEE models, multilevel models [34], and raw means. These methods represent common longitudinal statistical methods in clinical trials [42]. The estimation of change through all three GEE, mixed models, and raw scores was designed to clarify that the underlying function of symptom change could be identified when using various statistical models.

Under a linear pattern of symptom change, participants of any baseline symptoms would be expected to show a similar rate of improvement overall. That is, an average symptom change score that would be observed across individuals, irrespective of the severity of their symptoms at baseline [18]. In contrast, under a proportional pattern of symptom change, participants presenting with increased baseline symptom severity would likely show larger symptom change compared to those individuals with mild or moderate baseline symptoms [15].

To test the second hypothesis that distributions of pretreatment and posttreatment depression symptom levels would show evidence of positive skewness and kurtosis, the distributions of depression symptoms scores at both pretreatment and posttreatment were evaluated for evidence of skewness. In this step, if the dataset would present with statistically normal distribution of symptom scores at both time points, the symptom change over time would be considered as linear. In contrast, if symptoms changed as a proportional function from baseline, positive skewness should be observed, particularly at posttreatment, where individuals from various baseline symptoms would shift and concentrate around the symptom score band of minimal symptoms. Graphical and numerical explorations of pre-post score distributions were included.

To test the third hypothesis that statistical methods measuring symptom change as a proportional function would be associated with reduced measurement error and indicate greater statistical fit to real symptom data in treatment, the relative measurement accuracy of models that represent either linear or proportional symptom change were compared. Specifically, this step compared model fit statistics and the remaining unexplained (residual) variance associated with each function of change. Both mixed models and GEE models were run initially as models that assume change was linear, represented through models that specified a normal scale of the dependent variable and an identify link function. Following this, alternative statistical models were compared, which specified a gamma scale and a log link function; representing models that assumed change was proportional. Generally, the gamma scale is considered a suitable method for data showing signs of skewness and multiplicative change function [15]; however, the selection of the gamma scale does not imply that alternative multiplicative statistical methods (eg, negative binomial scale, Poisson scale, or zero inflated models) would be less effective.

Formulas emphasizing the difference in statistical notation between the multiplicative model (Equations 1.1-1.2) and the linear model (Equations 1.3-1.5) are presented in Figure 1. With more formal statistical notation, the multiplicative effect within the log link model is created when the intercept, β_0, or baseline symptoms, is multiplied by the treatment effect, β_{tj}, the estimate of exponential change following treatment (Equations 1.6-1.8 in Figure 1).

The suitability of either model type was evaluated through model fit statistics, generated using SAS 9.4 software. Specifically, the quasilikelihood under the independence model criterion (QIC) statistic [43] for GEE models, and Akaike information criterion (AIC) and Bayesian information criterion (BIC) for mixed effects models [44], compared between linear (additive) and generalized linear (proportional) models. Within all AIC, BIC, and QIC model fit estimates, relatively lower scores imply overall reduced variance, and overall increase measurement accuracy.

Figure 1. Equations 1.1-1.8.

Multiplicative model

(1.1) $Y_{ij} \sim Gamma\ (\mu_{ij}, \alpha)$

(1.2) $log\ (\mu_{ij}) = \beta 0 + \beta_{tj} + \epsilon_{ij}$

Linear additive model

(1.3) $Y_{ij} = \beta_0 + \beta_{tj} + \epsilon_{ij}$

(1.4) $\epsilon_{ij} \sim N\ (0, 1)$

(1.5) $i = 1,...,1098;\ j = 0,1$

$t_j = \{ 0\ (\text{time} = \text{pre-treatment});\ 1\ (\text{time} = \text{post treatment})$

β_0 is the random intercept at pre-treatment;

and β_{tj} is the treatment effect of change over time

Linear additive model

(1.6) $\hat{\mu}_{baseline} = e^{\hat{\beta} 0}$

(1.7) $\hat{\mu}_{postreatment} = e^{\hat{\beta} 0} * e^{\hat{\beta} t1}$

(1.8) $\hat{\mu}_{postreatment} = \hat{\mu}_{baseline} * e^{\hat{\beta} t1}$

In addition to model fit statistics, the measurement error associated with the assumption that symptom change was either a fixed average score, or a percentage improvement score, was compared. In this step, measurement error was created for each participant by comparing the predicted posttreatment score under each change assumption (eg, PHQ-9 change of 5 points or 50% from baseline) against a known participant outcome score at posttreatment. The difference between the expected symptom outcome and actual treatment outcome effectively represents measurement error under the two change assumptions, akin to residual scores and measurement error variance. The pattern of residuals created under either assumption of symptom change was explored in two ways. First, the total quantity of error variance under each function was compared. Second, measurement residuals were graphically explored under each function of symptom change by comparing the increase or decrease of residuals for individuals with different baseline symptom score.

Results

In the first step (operationalizing the first hypothesis that individuals with increased symptoms at baseline would also demonstrate increased rates of symptom change), the relationship between baseline symptom severity and the quantity of symptom change was explored graphically. Figure 2, illustrating PHQ-9 change as a linear function, and Figure 3, illustrating PHQ-9 change as a proportional change from baseline, both demonstrate the symptom change on the y-axis within each of the PHQ-9 baseline symptom bands (x-axis). In addition, the symptom change observed within the waitlist condition is included as a dotted trend line, illustrating the trend of nonspecific change in symptoms within each bands of symptom severity at baseline.

Figure 2 illustrates an increased rate of symptom change that was associated closely with increased baseline symptoms. In Figure 2, individuals with severe baseline symptoms were observed to reduce by as much as threefold compared to individuals with mild baseline symptoms (11.4 vs 3.7, respectively). In addition, participants with severe symptoms in the control group demonstrated a sizable reduction in symptoms even when treatment was not applied. This nonspecific symptom-related change was pronounced to the extent that individuals with severe baseline symptoms in the control group demonstrated higher symptom reduction than individuals with moderate symptoms in treatment (7 points vs 6 points, respectively). That is, as a linear effect, the nonspecific symptom change within the control condition was larger than the treatment-related symptom change of individuals with moderate symptoms.

Figure 3 illustrates the proportional percentage change of symptoms within each of the mild, moderate, moderately severe, and severe subgroups. The figure illustrates that as a proportional change, an average treatment-related change of 50% to 55% was observed across all subgroups of individuals who started with at least mild symptoms at baseline. Of note, the rate of proportional improvement in treatment (50%-55%) was greater than the nonspecific change experienced by individuals with severe baseline symptoms in the waitlist conditions (35%). That is, the measurement of change as a percentage change resulted in a clearer differentiation of treatment-specific and nonspecific change.

Figure 2. Measurement of mean treatment-related PHQ-9 symptom change per initial pretreatment symptom severity band; whiskers represent 95% CI s. Symptom change observed under control conditions indicated by a solid trend line.

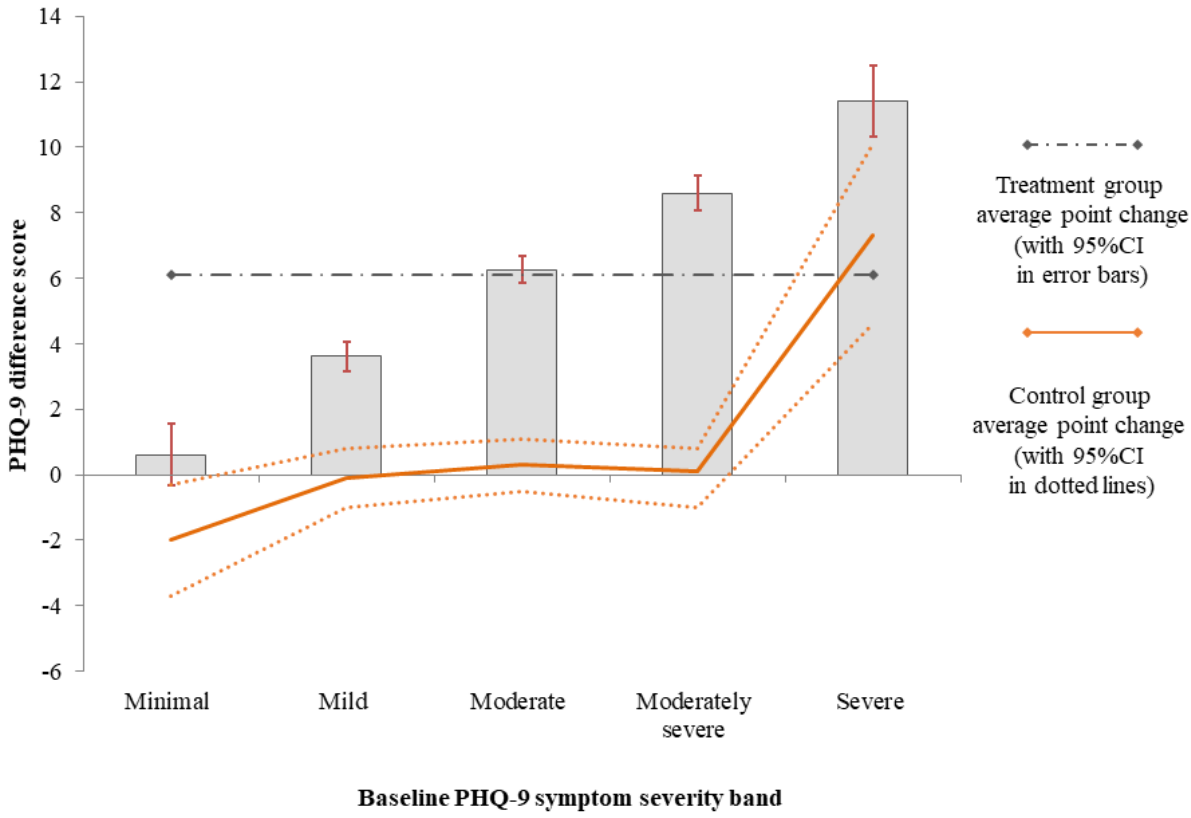


Figure 3. Measurement of mean treatment-related PHQ-9 symptom change as a proportional pattern of remission (52%); per initial pretreatment symptom severity; whiskers represent 95% CIs. Symptom change observed under control conditions indicated by a solid trend line.

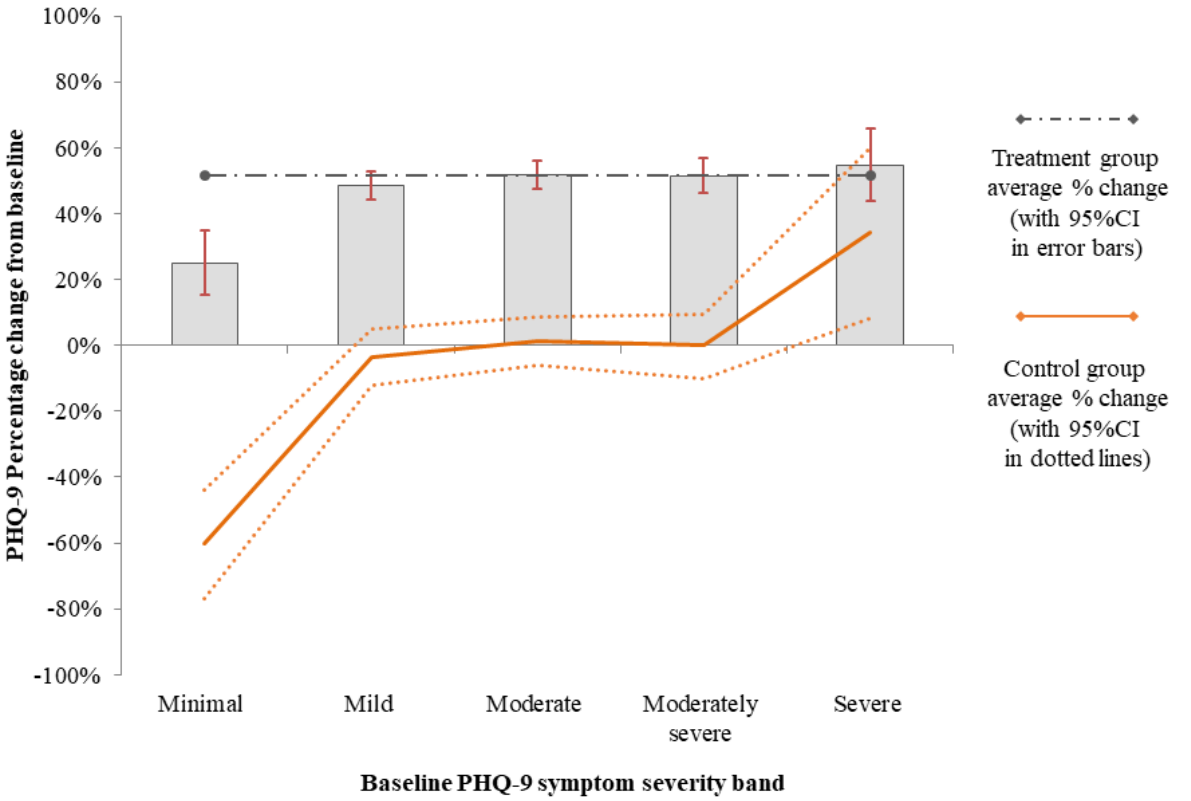


Table 2 includes the numerical descriptions of change for both the treatment and control conditions. Table 2 also includes effect sizes that were calculated within the treatment group as a whole and the effect size demonstrated by individuals in the mild, moderate, moderately severe, and severe bands of baseline symptoms. Individuals with mild depressive symptoms showed smaller effects (1.59) compared to individuals with more severe symptoms (3.9).

In a second step, the second hypothesis that distributions of pretreatment and posttreatment depression symptom levels would show evidence of positive skewness and kurtosis was operationalized with an exploration of the distribution of pretreatment and posttreatment symptom scores. Figure 4 illustrates the distribution of PHQ-9 symptom scores, both before and following treatment. These histograms illustrate a slight positive skewness of scores at pretreatment, with fewer individuals presenting within the severely symptomatic band as compared to the mild and moderate bands. In contrast, at posttreatment, increasing positive skewness was observed, where most individuals who reduced their symptoms became concentrated within the mild to minimal symptom ranges. The numerical estimates of the skewness are collated in Table 3.

Taken together, both numerically and graphically, the distributions of symptom scores demonstrated significant positive skewness that increased at posttreatment.

In a third step, the third hypothesis that statistical methods measuring symptom change as a proportional function would be associated with reduced measurement error and indicate greater statistical fit to real symptom data in treatment was operationalized, seeking to explore the model fit of the linear and the multiplicative statistical models of symptom change. Table 4 collates the goodness-of-fit statistics from models that specified either a proportional or linear function of change.

In Table 4, models that specified a proportional function of symptom change demonstrated a several-fold improvement in the model fit statistics within both the GEE and mixed models, including reduced QIC statistics, reduced AIC, and reduced BIC estimates. Table 4 also collated the measurement error associated with the prediction that change occurred as a linear change of six points, or as a percentage improvement (52% reduction from baseline). A notable reduction in the total estimate of PHQ-9 error variance was evident when a proportional function of change was assumed (σ^2=16.716 vs σ^2=24.122). This result demonstrated that by characterizing change as a proportional function, the measurement error and remaining unknown individual variation reduced by more than 30%.

Table 2. Rates of change of nine-item Patient Health Questionnaire (PHQ-9) scores associated with linear and proportional change functions; estimates per initial baseline symptom subgroups.

PHQ-9 and change functions	Initial symptom severity					Total
	Minimal (n=72)	Mild (n=345)	Moderate (n=381)	Moderately severe (n=244)	Severe (n=56)	Overall sample (treatment) scores
Observed PHQ-9, mean (SD)						
Pretreatment	2.83 (1.25)	7.32 (1.33)	12.07 (1.40)	16.67 (1.41)	20.86 (0.84)	11.41 (4.79)
Posttreatment	2.22 (2.61)	3.71 (3.3)	5.81 (3.92)	8.07 (5.41)	9.45 (4.99)	5.59 (4.57)
GEE[a] (95% CI)[b]						
Additive change estimate	0.61 (–0.30 to 1.18)	3.66 (3.30 to 4.02)	6.22 (5.82 to 6.62)	8.66 (7.98 to 9.34)	11.43 (10.14 to 12.73)	6.00 (5.71 to 6.28)
Percent proportional change estimate	21% (–1 to 39)	50% (45 to 54)	52% (48 to 55)	52% (48 to 56)	55% (48 to 61)	52 (50 to 54)
Effect size, Cohen *d* (95% CI)	0.32 (0.01 to 0.63)	1.59 (1.43 to 1.74)	2.34 (2.19 to 2.49)	2.54 (2.33 to 2.74)	3.90 (3.45 to 4.36)	1.27 (1.21 to 1.34)
Control group						
Change[c] (95% CI)[b]	–2 (–27 to –1.24)	–0.1 (–0.76 to 0.53)	0.29 (–0.68 to 1.28)	0.48 (–1.01 to 1.15)	7.37 (5.14 to 9.51)	0.68 (–0.37 to 0.16)
Percent proportional change estimate, GEE (95% CI)[b]	–61 (–78 to –44)	–4 (–12 to 5)	1 (–6 to 9)	0 (–10 to 10)	34 (8 to 60)	0% (–1 to 1)

[a]GEE: generalized estimated equation.

[b]Confidence intervals based on modeled marginal means.

[c]Control group change is nonspecific effect.

Figure 4. Dispersion of symptom scores (nine-item Patient Health Questionnaire, PHQ-9) at pretreatment (in light bars) and posttreatment scores (in dark bars). The dotted trend lines are indicative of the shape of each distribution.

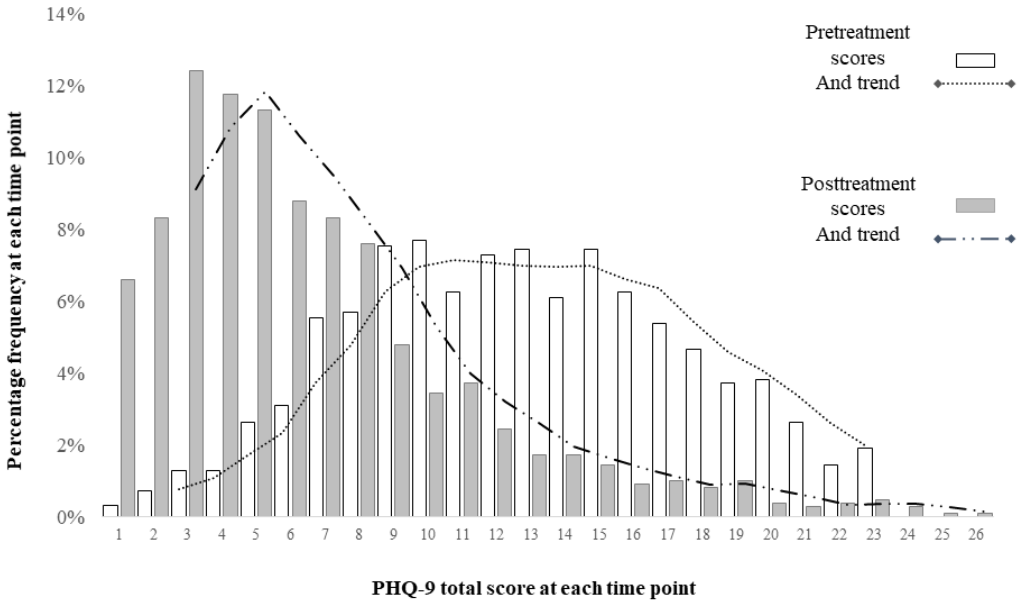


Table 3. Symptom score distributions statistics

Sample and time point	Skewness (SE)	Baseline symptoms, mean (SD)	Effect size, Cohen *d* (95% CI)
Treatment sample (n=1098)			1.27 (1.21 to 1.34)
Pretreatment	0.271 (0.071)[a]	11.73 (4.83)	
Posttreatment	1.359 (0.076)[a]	5.60 (4.58)	
Control sample depression (n=96)			–0.04 (–0.24 to 0.16)
Pretreatment	0.178 (0.109)	10.91 (4.53)	
Posttreatment	0.228 (0.109)	11.00 (5.04)	

[a]Statistical significance beyond .05 alpha on a Shapiro-Wilk test for distribution normality; significance is indicative that normal distribution is not supported within the observed sample.

Table 4. Model fit statistics and dispersion of model residuals for the treatment sample (n=1098). Model fit criterion was derived from SAS software, version 9.3.

Method of change specified	QIC[a,b] (GEE[c] model)	AIC[d,b] (Mixed)	BIC[e,b] (Mixed)	Total variance (PHQ-9 σ^2)
Linear (normal scale)	52457.6	14059.8	14071.3	16.716
Proportional (gamma scale)	2020.5	4041.8	4053.3	24.122

[a]QIC: quasilikelihood under the independence model criterion.

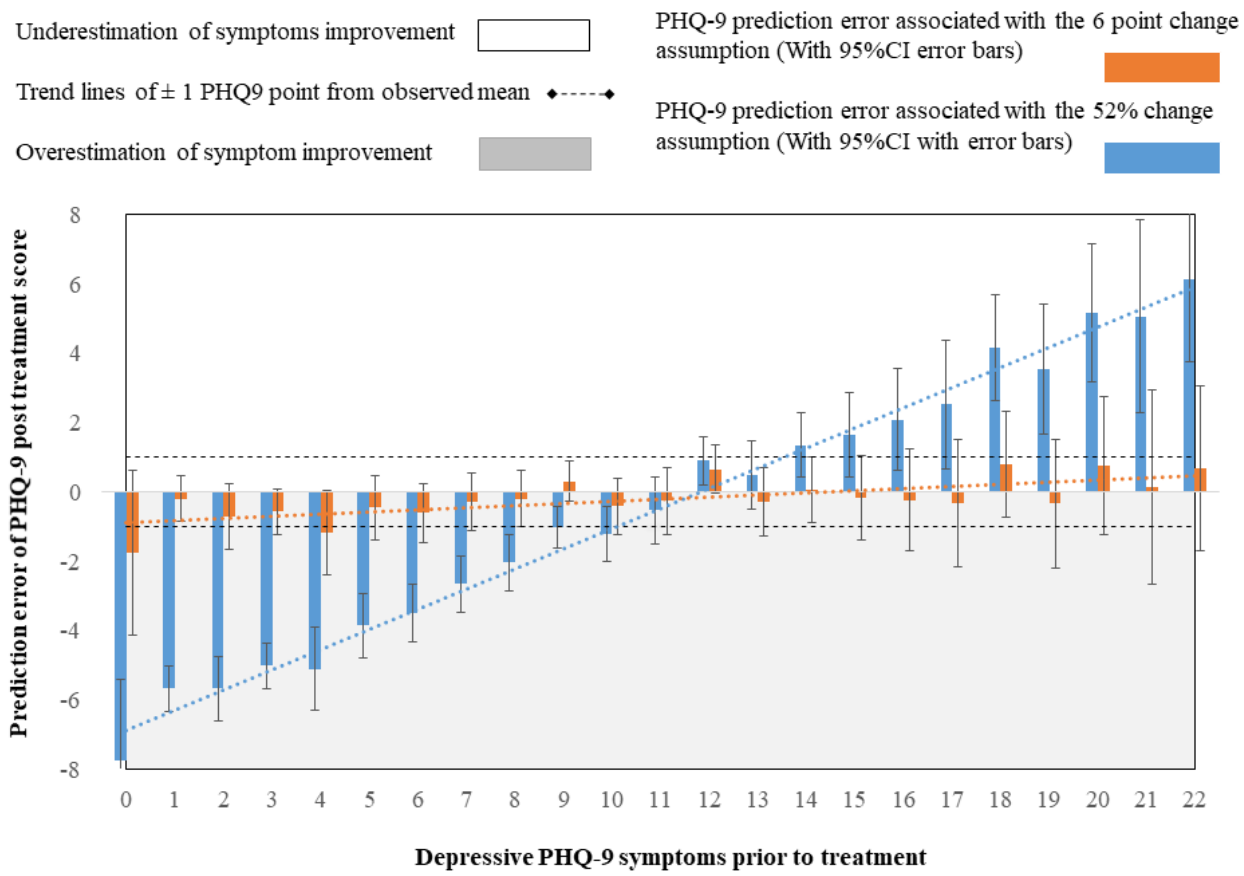
[b]Confidence intervals based on the multiplicative longitudinal GEE model specified in the analytical plan.

[c]GEE: generalized estimated equation.

[d]AIC: Akaike information criterion.

[e]BIC: Bayesian information criterion.

Figure 5. PHQ-9 estimation error (residual) following fixed (linear) and relative (proportional) change assumption.



The measurement error associated with either assumption that change was linear (6 points) or proportional (52%) were graphically explored. Figure 5 illustrates the residual error (y-axis) across individuals who started treatment with different baseline symptoms (x-axis). In the figure, individuals with mild and severe baseline symptoms can be observed to substantially underestimate or overestimate the rate of symptom change when linear change (6 points) was predicted. In contrast, when change was predicted to be proportional (52%), baseline symptoms no longer associated with the rate measurement error. Further, under the proportional assumption, the predicted symptom outcome could be accurately predicted within a single point across individuals with different baselines (marked with dots horizontal lines). In contrast, under the linear assumption, the prediction of symptom outcome become systematically erroneous with baseline severity (a range of up to 16 points between mild and severe).

Discussion

This study aimed to investigate the statistical characteristic of symptom change in treatment and compare different ways to measure and interpret symptom change. Using a Web-based psychotherapy sample (n=1098), as well as a waitlist control condition (n=96), the statistical characterization of depressive symptom change (PHQ-9) was explored in three steps, corresponding to three proposed hypotheses.

Testing of the first hypothesis demonstrated support for the characterization of symptom change as a proportional function through a clear association between symptom severity at baseline and the rate of change. In contrast, as a proportional estimate of change, individuals in treatment demonstrated a consistent rate of proportional symptom change within all subgroups with mild, moderate, moderately severe, and severe baseline symptom (50%-55%). Critically, the dependency between symptom change and baseline symptom severity was also observed in the waitlist condition, with mild and severe participants changing proportionally in their symptoms even when treatment was not applied. Testing of the second and third hypotheses also illustrated support for the characterization of symptom change as proportional function, with symptom score distributions presenting with positive skewness, particularly following treatment (H2). Similarly, increased model fit, and reduced measurement error was observed when the treatment sample was statistically modeled with an underlying proportional function of change (H3).

The analyses within this study are novel in that they characterize the function of depressive symptom change and compare different statistical methods for measuring and interpreting symptom change within treatment as well as nontreatment conditions. The findings suggest that common psychotherapy symptom scales (eg, PHQ-9) are impacted by a feature of natural bounding at minimal symptoms, which is the suspected culprit for the resulting (1) nonnormal distributions at posttreatment,

(2) the dependency between baseline symptoms and rate of change, and (3) the improved model fit for techniques that assume longitudinal change is proportional to baseline.

These findings raise two potentially critical implications for the ability to measure and interpret psychotherapy change in combination with symptom scales. First, the inappropriate use of linear statistics, such as Cohen *d*, when change is proportional would lead to artificially higher estimates of clinical efficacy, both in treatment and in control conditions. For example, in this study, individuals with severe baseline symptoms demonstrated effect sizes that increased by nearly threefold (3.9) when compared to individuals with mild symptoms (1.6), even when the same treatment was applied. This is problematic because linear estimates of change such as Cohen *d* are strongly associated with baseline severity and not with quality or the effectiveness of treatment. This finding is broadly consistent with the data within previous psychotherapy studies showing increased effect sizes with samples of increased symptoms, even when similar treatments are applied [20,29,32].

Second, these findings support a well-established statistical idea posing that the selection of a statistical analysis must match the characteristics of the dataset in order to arrive at valid and accurate statistical measurement, interpretation, and conclusions [4,45]. In this context of depressive symptom scales, the use of proportional statistical analyses resulted in (1) improved statistical modeling of treatment effects, (2) an improved ability to determine what a treatment effect is (50%-55%) and what a nontreatment effect is (35%), as well as for (3) establishing a clinical effect that is robust across individuals with various baseline symptoms (50%-55%). The measurement and interpretation of change as proportional improvement from baseline can also be concretely and easily interpreted as an estimate of change (eg, percentage improvement). Further, in the context of treatment, percentage improvement and percentage change estimates seem to reflect the ideal of treatment (reducing symptoms to minimal) [1,9]. For these reasons, measuring and interpreting change as a fundamentally proportional function can hold critical implications for clinical research that is reliant on accurate and interpretable measurement. For example, researchers seeking to identify clinical moderators, compare between treatments, estimate cost-effectiveness, or classify individual effects are likely to be positively impacted with a suitable choice of analytics that capture the underlying statistical function of change [36,37].

Although the measurement and interpretation of symptom change as a proportional change show promise to increase the accuracy and interpretability of clinical change, several statistical and clinical limitations should be considered about the results of this study. Primarily, the results of this study should be considered as (1) preliminary, (2) specific to a symptom scale of depressive symptoms (PHQ-9), and (3) specific to one kind of treatment model (the Macquarie University online model). Specifically, albeit the strengths of this study as an exploration of change within a large and standardized sample, it is unclear to what extent the 50% to 55% symptom change is specific to this treatment model and to the PHQ-9 scale.

To address these limitations, statistical replication is needed across different symptom scales and treatment models. Specifically, the characterization of symptom change must be observed within other psychotherapy treatment models before more generalizable comments can be made about symptom change and measurement. Future similar studies seeking to characterize and compare symptom change and measurement models could determine to what extent the proportional change pattern generalizes as a measurement principle, across different treatment models and across different symptom scales. In addition, future studies seeking to research this pattern of change could also attempt to compile a meta-analytical characterization of proportional and linear change across different scales and treatment models.

Further, it is important to consider that measurement and interpretation of symptom change as a proportional function is at odds with the widely accepted use of linear statistics in psychotherapy. From one point of view, linear statistics, such as Cohen *d*, are successful as an established measurement standard that can be used to compare change estimates between trials and across clinical instruments [2]. This use of effect sizes has resulted in both enormous amounts of aggregated evidence about the effects of psychotherapy [22] and, for this reason, it is understandable clinical researchers would continue to use this standard for measuring and interpreting symptom change. However, should symptom change occur as a proportional function, the measurement and interpretation of treatment-related change would substantially improve by matching appropriate statistical analysis to the characteristics of the function of symptom change [15,45,46]. A possible solution to this dilemma would be to report both the effect size and percentage estimates of change side by side. In this way, the change that occurs in treatment can be more accurately reported, evaluated, and compared between trials.

Finally, this study does not weigh whether the change rate of 50% to 55% could be evaluated as the same treatment-related effect across individuals with severe or mild baseline symptoms. For example, a symptom reduction demonstrated by individuals with severe baseline symptoms could be interpreted as a more substantive clinical effect than an equivalent symptom reduction achieved with individuals with mild or moderate symptoms [47]. To address these limitations, additional research into the experience of individuals in treatment could determine whether individuals with different baseline symptoms consider the proportional remission pattern an equally satisfactory treatment outcome. For example, Zimmerman and colleagues [48] consider the measurement of patient functionality, positive mental health, and optimism alongside the reduction in depressive symptoms. These additional measures could verify and elaborate on the experience of individuals in treatment and nontreatment conditions, within various symptom bands, shedding more light on the universality or segmentation of the 50% to 55% improvement effect.

In summary, this study aimed to explore the underlying pattern of symptom change and compare different methods for measuring and interpreting depressive symptom change that follows treatment (Web-based psychotherapy). This study has combined evidence of increased rate of change with increased

baseline symptoms (hypothesis 1), score distributions that become increasingly skewed following treatment (hypothesis 2), and increased measurement accuracy achieved by statistical methods that assume change is proportional (hypothesis 3) to suggest that the fundamental function of symptom change is proportional. The promise of matching these characteristics of proportional symptom change to a suitable statistical analysis is important for all (1) statistical modeling and the prediction of treatment effects, (2) an improved ability to differentiate treatment and nonspecific symptom change, as well as for (3) determining an estimate of treatment-related change that will not sway with increased baseline symptoms. Replication of these preliminary findings are essential within additional depressive symptom scales, other types of psychological conditions, and across different treatment modalities.

Acknowledgments

The authors would like to acknowledge Monique Crane, Pery Karin, and the team of reviewers for their helpful and meticulous feedback.

Conflicts of Interest

None declared.

References

1. Kroenke K, Monahan PO, Kean J. Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. J Clin Epidemiol 2015 Sep;68(9):1085-1092 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.03.023] [Medline: 25962972]
2. Spring B. Evidence-based practice in clinical psychology: what it is, why it matters; what you need to know. J Clin Psychol 2007 Jul;63(7):611-631. [doi: 10.1002/jclp.20373] [Medline: 17551934]
3. Wise E. Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. J Pers Assess 2004 Feb;82(1):50-59. [doi: 10.1207/s15327752jpa8201_10] [Medline: 14979834]
4. Flay BR, Biglan A, Boruch RF, Castro FG, Gottfredson D, Kellam S, et al. Standards of evidence: criteria for efficacy, effectiveness and dissemination. Prev Sci 2005 May 16;6(3):151-175. [doi: 10.1007/s11121-005-5553-y] [Medline: 28116558]
5. Gottfredson DC, Cook TD, Gardner FE, Gorman-Smith D, Howe GW, Sandler IN, et al. Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: next generation. Prev Sci 2015 Apr 7;16(7):893-926. [doi: 10.1007/s11121-015-0555-x] [Medline: 25846268]
6. Laurenceau J, Hayes AM, Feldman GC. Some methodological and statistical issues in the study of change processes in psychotherapy. Clin Psychol Rev 2007 Jul;27(6):682-695 [FREE Full text] [doi: 10.1016/j.cpr.2007.01.007] [Medline: 17328996]
7. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot Clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
8. Gunn J, Elliott P, Densley K, Middleton A, Ambresin G, Dowrick C, et al. A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. J Affect Disord 2013 Jun;148(2-3):338-346. [doi: 10.1016/j.jad.2012.12.021] [Medline: 23375580]
9. Sobocki P, Ekman M, Agren H, Runeson B, Jönsson B. The mission is remission: health economic consequences of achieving full remission with antidepressant treatment for depression. Int J Clin Pract 2006 Jul;60(7):791-798. [doi: 10.1111/j.1742-1241.2006.00997.x] [Medline: 16846399]
10. Gyani A, Shafran R, Layard R, Clark DM. Enhancing recovery rates: lessons from year one of IAPT. Behav Res Ther 2013 Sep;51(9):597-606 [FREE Full text] [doi: 10.1016/j.brat.2013.06.004] [Medline: 23872702]
11. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. J R Soc Med 2016 Feb;109(2):67-77. [doi: 10.1177/0141076815625599] [Medline: 26880653]
12. Choi S, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychol Assess 2014 Jun;26(2):513-527 [FREE Full text] [doi: 10.1037/a0035768] [Medline: 24548149]
13. Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. J Anxiety Disord 2014 Jan;28(1):88-96 [FREE Full text] [doi: 10.1016/j.janxdis.2013.11.006] [Medline: 24508596]
14. Snyder C, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012 Oct;21(8):1305-1314 [FREE Full text] [doi: 10.1007/s11136-011-0054-x] [Medline: 22048932]

15. Baldwin SA, Fellingham GW, Baldwin AS. Statistical models for multilevel skewed physical activity data in health research and behavioral medicine. Health Psychology 2016;35(6):552-562. [doi: 10.1037/hea0000292] [Medline: 26881287]
16. Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials 2007 Nov 26;8(1):39 [FREE Full text] [doi: 10.1186/1745-6215-8-39] [Medline: 18039365]
17. Horn SD, Gassaway J. Practice-based evidence study design for comparative effectiveness research. Med Care 2007 Oct;45(10 Supl 2):S50-S57. [doi: 10.1097/MLR.0b013e318070c07b] [Medline: 17909384]
18. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
19. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
20. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013 Feb 26;346(feb26 2):f540-f540. [doi: 10.1136/bmj.f540] [Medline: 23444423]
21. Clark DM. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. Int Rev Psychiatry 2011 Aug;23(4):318-327 [FREE Full text] [doi: 10.3109/09540261.2011.606803] [Medline: 22026487]
22. Newby J, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
23. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND Statement. Am J Public Health 2004 Mar;94(3):361-366. [doi: 10.2105/AJPH.94.3.361]
24. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg 2014 Dec;12(12):1495-1499 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
25. Hiller W, Schindler AC, Lambert MJ. Defining response and remission in psychotherapy research: a comparison of the RCI and the method of percent improvement. Psychother Res 2012 Jan;22(1):1-11. [doi: 10.1080/10503307.2011.616237] [Medline: 21943215]
26. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. J Affect Disord 2010 Dec;127(1-3):122-129 [FREE Full text] [doi: 10.1016/j.jad.2010.04.030] [Medline: 20569992]
27. Ellis PD. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge, UK: Cambridge University Press; 2010.
28. Boettcher J, Hasselrot J, Sund E, Andersson G, Carlbring P. Combining attention training with internet-based cognitive-behavioural self-help for social anxiety: a randomised controlled trial. Cogn Behav Therapy 2013 Jul 30;43(1):34-48. [doi: 10.1080/16506073.2013.809141] [Medline: 23898817]
29. Kroenke K, Spitzer RL, Williams JB. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x]
30. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317. [doi: 10.7326/0003-4819-146-5-200703060-00004]
31. Kessler R, Andrews G, Colpe L, Hiripi E, Mroczek D, Normand S, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002;32(6):959-976. [doi: 10.1017/S0033291702006074]
32. Driessen E, Cuijpers P, Hollon SD, Dekker JJ. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. J Consult Clin Psych 2010;78(5):668-680. [doi: 10.1037/a0020570] [Medline: 20873902]
33. Thase M, Simons AD, Cahalane J, McGeary J, Harden T. Severity of depression and response to cognitive behavior therapy. Am J Psychiatry 1991 Jun;148(6):784-789 [FREE Full text] [doi: 10.1176/ajp.148.6.784] [Medline: 2035722]
34. Fitzmaurice GM. In: Laird NM, Ware JH, editors. Applied Longitudinal Analysis. Philadelphia, PA: John Wiley & Sons; 2012.
35. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986 Apr;73(1):13. [doi: 10.2307/2336267]
36. Castellani B, Rajaram R, Gunn J, Griffiths F. Cases, clusters, densities: Modeling the nonlinear dynamics of complex health trajectories. Complexity 2015 Sep 25;21(S1):160-180. [doi: 10.1002/cplx.21728] [Medline: 25855820]
37. Panagiotakopoulos TC, Lyras DP, Livaditis M, Sgarbas KN, Anastassopoulos GC, Lymberopoulos DK. A contextual data mining approach toward assisting the treatment of anxiety disorders. IEEE Trans Inf Technol Biomed 2010 May;14(3):567-581. [doi: 10.1109/TITB.2009.2038905] [Medline: 20071265]
38. Pocock S, Clayton TC, Stone GW. Challenging issues in clinical trial design: part 4 of a 4-part series on statistics for clinical trials. J Am Coll Cardiol 2015 Dec 29;66(25):2886-2898 [FREE Full text] [doi: 10.1016/j.jacc.2015.10.051] [Medline: 26718676]

39. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: A randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
40. Dear B, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
41. eCentreClinic. URL: https://www.ecentreclinic.org/ [accessed 2018-06-07] [WebCite Cache ID 700kbKH6d]
42. Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Lippman SA, Jewell N, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology 2010 Jul;21(4):467-474. [doi: 10.1097/EDE.0b013e3181caeb90] [Medline: 20220526]
43. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57(1):120-125 [FREE Full text] [doi: 10.1111/j.0006-341X.2001.00120.x]
44. Akaike H. Information theory and an extension of the maximum likelihood principle. 1973 Presented at: 2nd International Symposium on Information Theory; Sep 2-8, 1971; Tsahkadsor, Armenia, USSR.
45. Field AP, Wilcox RR. Robust statistical methods: a primer for clinical psychology and experimental psychopathology researchers. Behav Res Ther 2017 Nov;98:19-38. [doi: 10.1016/j.brat.2017.05.013] [Medline: 28577757]
46. Verkuilen J, Smithson M. Mixed and mixture regression models for continuous bounded responses using the beta distribution. J Educ Behav Stat 2016 Aug 26;37(1):82-113. [doi: 10.3102/1076998610396895]
47. Judd LL, Schettler PJ, Rush AJ, Coryell WH, Fiedorowicz JG, Solomon DA. A new empirical definition of major depressive episode recovery and its positive impact on future course of illness. J Clin Psychiatry 2016 Aug;77(8):1065-1073. [doi: 10.4088/JCP.15m09918] [Medline: 26580150]
48. Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Attiullah N, Boerescu D. How should remission from depression be defined? The depressed patient's perspective. Am J Psychiatry 2006 Jan;163(1):148-150. [doi: 10.1176/appi.ajp.163.1.148] [Medline: 16390903]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
GEE: generalized estimated equation
ICBT: Internet-delivered cognitive behavioral therapy
QIC: quasilikelihood under the independence model criterion

Edited by J Lipschitz; submitted 23.02.18; peer-reviewed by M Subotic-Kerry, BT Tulbure; comments to author 13.04.18; revised version received 03.05.18; accepted 07.05.18; published 12.07.18.

Please cite as:
Karin E, Dear BF, Heller GZ, Gandy M, Titov N
Measurement of Symptom Change Following Web-Based Psychotherapy: Statistical Characteristics and Analytical Methods for Measuring and Interpreting Change
JMIR Ment Health 2018;5(3):e10200
URL: http://mental.jmir.org/2018/3/e10200/
doi:10.2196/10200
PMID:30001999

Original Paper

An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial

Shinichiro Suganuma[1], PhD; Daisuke Sakamoto[2], PhD; Haruhiko Shimoyama[1], PhD

[1]Department of Clinical Psychology, Graduate School of Education, The University of Tokyo, Tokyo, Japan

[2]Division of Computer Science and Information Technology, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan

Corresponding Author:
Shinichiro Suganuma, PhD
Department of Clinical Psychology
Graduate School of Education
The University of Tokyo
Tokyo,
Japan
Phone: 81 3 5841 8068
Fax: 81 3 5841 3949
Email: sgnm.shin@gmail.com

Abstract

Background: Recent years have seen an increase in the use of internet-based cognitive behavioral therapy in the area of mental health. Although lower effectiveness and higher dropout rates of unguided than those of guided internet-based cognitive behavioral therapy remain critical issues, not incurring ongoing human clinical resources makes it highly advantageous.

Objective: Current research in psychotherapy, which acknowledges the importance of therapeutic alliance, aims to evaluate the feasibility and acceptability, in terms of mental health, of an application that is embodied with a conversational agent. This application was enabled for use as an internet-based cognitive behavioral therapy preventative mental health measure.

Methods: Analysis of the data from the 191 participants of the experimental group with a mean age of 38.07 (SD 10.75) years and the 263 participants of the control group with a mean age of 38.05 (SD 13.45) years using a 2-way factorial analysis of variance (group × time) was performed.

Results: There was a significant main effect (P=.02) and interaction for time on the variable of positive mental health (P=.02), and for the treatment group, a significant simple main effect was also found (P=.002). In addition, there was a significant main effect (P=.02) and interaction for time on the variable of negative mental health (P=.005), and for the treatment group, a significant simple main effect was also found (P=.001).

Conclusions: This research can be seen to represent a certain level of evidence for the mental health application developed herein, indicating empirically that internet-based cognitive behavioral therapy with the embodied conversational agent can be used in mental health care. In the pilot trial, given the issues related to feasibility and acceptability, it is necessary to pursue higher quality evidence while continuing to further improve the application, based on the findings of the current research.

(JMIR Ment Health 2018;5(3):e10454) doi:10.2196/10454

KEYWORDS

embodied conversational agent; cognitive behavioral therapy; psychological distress; mental well - being; artificial intelligence technology

Introduction

Current State of Mental Health Care Utilizing Information and Communication Technologies, and its Significance

In the current era, the use of biological, psychological, and social models [1] to understand and support mental health issues is common, and of these, cognitive behavioral therapy (CBT) is one of the major means of providing psychological understanding and support. Owing to its confirmed efficacy across a range of mental health care-related fields, from changing daily habits to specialist interventions for mental illness, CBT has become a psychological approach used worldwide and has been delivered in various forms to date, including face-to-face, group therapy, and books (refer to Westbrook, Kennerley, and Kirk [2]). Meanwhile, there has long been a problem with patients not visiting a clinician or not discussing their concerns, even when there is a mental health issue, and research into the service gap, that is, the difference between the need for and uptake of mental health services, has been ongoing [3]. Underlying this is the dual problem of support providers and practitioners being unable to reach those in need of services, as well as that of service users being unable to access services. The former involves issues of privacy and cost concerning service provision, while the latter deals with the stigma surrounding psychiatry and mental illness, as well as preferences regarding methods of accessing support.

In recent years, internet-based CBT (ICBT) and computerized CBT have increasingly been used as a means to fill the service gap and resolve various problems related to mental health, such as bipolar disorder [4,5], anxiety disorder [6], depression [7], treatment adherence [8], and common mental health problems [9]. According to reviews published by Andersson [10,11], therapist-guided ICBT is standard, and while its efficacy in the three characteristic areas of depression, anxiety, and physical symptoms has been shown to be almost equivalent to that of face-to-face CBT, the issues of lower efficacy and higher dropout rates of unguided ICBT than those of guided ICBT have been raised. In 2016, in a systematic review of depression-related self-help smartphone apps by Huguet et al [12], it became apparent that there are no suitable, evidenced-based CBT and behavioral activation (BA) apps, that is, no unguided ICBT apps are available, despite the large societal demand for and number of apps available.

Potential of Mental Health Care With an Embodied Conversational Agent

However, given that service provision cost was one of the factors motivating the original use of ICBT, unguided ICBT is a very attractive option because ongoing human clinical resources are unnecessary. When considered in terms of underlying clinical issues, the therapeutic alliance issue could conceivably underlie the problem of ICBT's low efficacy and high dropout rate. The therapeutic alliance has been considered an integral element in not just face-to-face CBT, but in all forms of psychotherapy, in terms of its role as a common factor in the efficacy of treatment [13]. The therapeutic alliance, as formed in unguided ICBT, is noted to be lacking in terms of elements of "development" and "maintenance" [14]. As such, artificial intelligence (AI) technology, especially embodied conversational agent (ECA), is conceivable as an effective means of using technology to overcome the paucity of therapeutic alliance-developing elements in unguided ICBT. In fact, its efficacy in reducing symptoms of depression has been reported by Fitzpatrick et al [15] in their research, utilizing a completely automated text-based response agent based on CBT principles. By offering pseudo-dialog experiences of freestyle dialog with the agent, the therapist-role agent in particular allows for the development of the therapeutic alliance even in unguided ICBT, and thus could contribute to increased efficacy and reduced dropout rates.

ECAs are electronic agents that have some type of embodiment and communicate messages to users [16]. ECAs are associated with the popular term "chatbot." Usage of agent-based technology for mental health care is still in its infancy. According to a scoping review for ECA applications in clinical psychology, more than half of their studies were focused on autism-related treatments and most applications were still under development and pilot phases [17]. Furthermore, there were other applications related to detecting and preventing suicidal behavior [18], changing of stigmatizing attitudes [19], and mental health interview [20]. Moreover, for psychological interventions, certain famous mental health chatbots existed such as "Woebot" [21] and "Tess" [22]. Historically, virtual affective agents were utilized in serious games for health care [23]. However, most of these are limited to narrow contexts and do not enable complex natural-language interactions [23]. Thus, currently, there is insufficient information regarding the combined use of an ECA that is capable of text-based interactions with traditional ICBT and its efficacy.

Against this background, the current research involved an evaluation of the feasibility and acceptability, in terms of mental health, of an app with an ECA that is capable of text-based freestyle dialog enabled for use in ICBT preventative mental health. In doing so, the investigation was conducted using a nonrandomized comparison study, in light of research reporting the general ineffectiveness and high dropout rate of unguided ICBT [10,11], as well as the current general lack of awareness regarding agent use in ICBT. While nonrandomized prospective studies (NPSs), such as a nonrandomized comparison study, are inferior to randomized comparison studies in terms of quality of evidence, they offer benefits such as significant low risk of discontinuation [24] and also allow the most economic use of resources. Moreover, in accordance with the aims of preventative mental health measures, positive and negative mental health effects indices were used as general mental health indicators.

Methods

Unguided Internet-Based Cognitive Behavioral Therapy Method

We use the "SABORI" as an ICBT application, which is a self-care application developed by the Laboratory of third author (The University of Tokyo), based on CBT and BA principles, for the purpose of preventing mental health issues. It is a Web-based unguided ICBT application available for use on a

smartphone, tablet, or computer browsers for company employees, university students, and housewives. SABORI users engage in self-monitoring by answering questions about changes in their daily mood and physical condition, subsequently receiving feedback and behavioral suggestions relevant to their responses. All user interactions in answering questions were only for the user to choose options; text-based interaction was not available. Improving one's own mental health state thus becomes intentional, through the promotion of self-monitoring and behavioral activation [25]. In this work, we improved and upgraded the interactive contents on the SABORI application to achieve better adherence. The two major reasons are as follows. First, the structure of SABORI was suited to adding a conversation-style dialog capability, because of the format of a user answering questions of the agent in the application. Second, since results from the preliminary study without a control group suggested improvement in mood due to behavioral activation for the depression group from one month of use [25], it was considered that addition of a freestyle dialog capability would be unlikely to adversely affect users. In this study, we upgraded SABORI by adding an agent-based freestyle dialog capability to the behavior suggestion section of SABORI and investigated its psychological effects. Multimedia Appendices 1 and 2 show screenshots of examples of freestyle dialog during monitoring and behavioral suggestions.

Development of the Conversation-Style Dialog Capability on the Internet-Based Cognitive Behavioral Therapy Application

First, the authors sought and gained ethics approval from our University Ethics Review Committee for the entire research plan, including this study. The majority of the development and implementation was conducted by two university lecturers specializing in clinical psychology, one information engineering specialist, and two specialists from a dialog system development company. The dialog system used in this research is an AI technology system, which includes multi-agent system as an AI engine, constructed from numerous agents with individual dictionaries or rules defined for each domain. The basic structure of conversation is that first the agent asks the user a behavior suggestion-related question and then responds to the user's input, and following this sequence, it transitions naturally to SABORI's existing behavior suggestion (Multimedia Appendices 1 and 2). The ontology for the system to classify user input and the creation of the behavior-related questions were constructed from September to December 2016. Basic knowledge for conversation was also collected from internet (eg, Wikipedia).

Human-in-the-Loop Improvement of the Freestyle Dialog Feature

In March 2017, a preliminary study was conducted to test and improve the freestyle dialog capability. The freestyle dialog capability was revised to form natural conversation patterns by having 10 clinical psychology students use it for one-month, allowing it to learn vocabulary and phrases not initially predicted and by altering the system's behavior suggestion method. In addition, during this period, 1162 cases of user input were gathered, and the system's response success rate was 66.52% at this stage. Moreover, the response success rate was determined as the total number of input cases excluding those that it was unable to respond to.

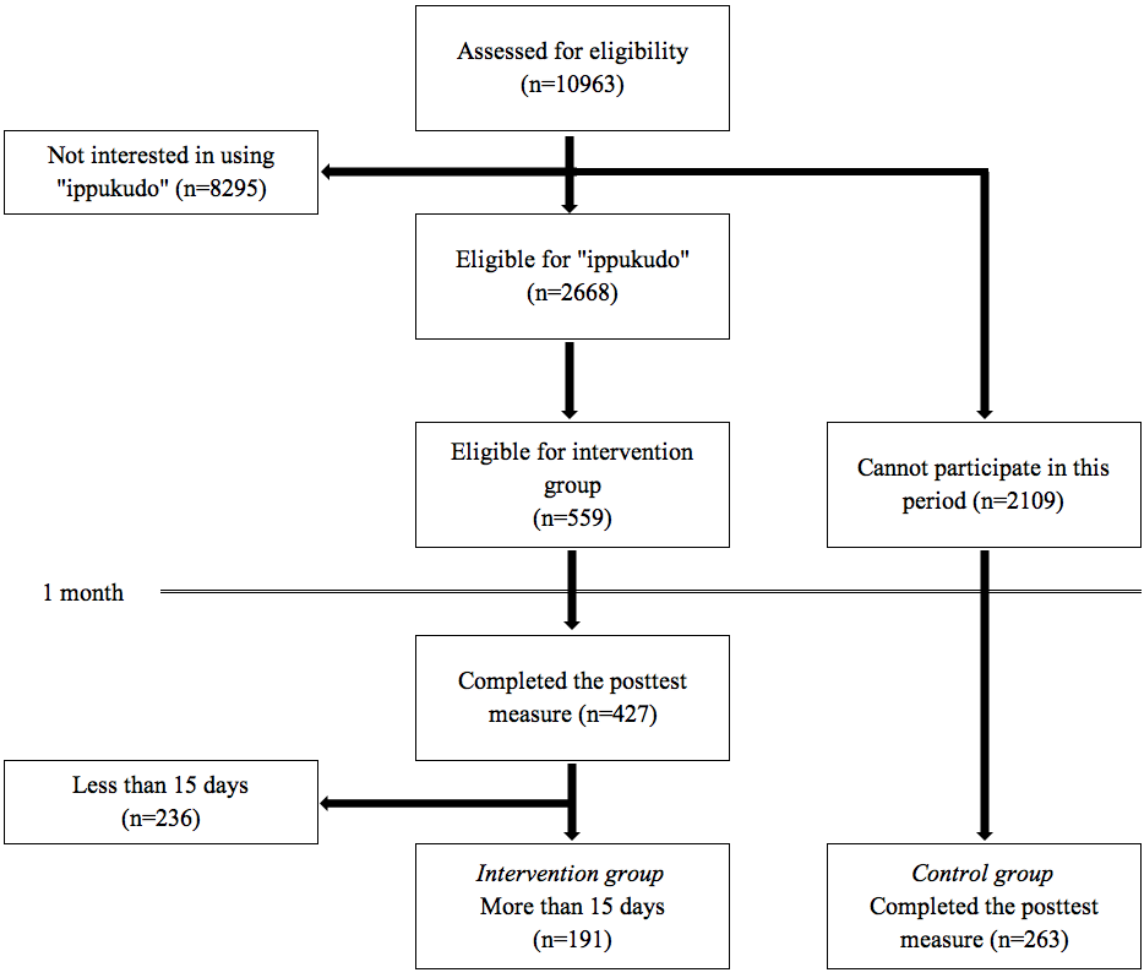
Furthermore, in the preliminary study, when the behavioral suggestion was provided using the user's unadulterated input, the uniqueness of the pattern was conspicuous to continuing users, and while the system was aware of this, from the user's perspective, it appeared that no change had occurred, thus creating the impression in many users that the system was unaware. Therefore, in consideration of the differences between the system's awareness rates and the user's perceived system awareness rates, official names were used where possible, and responses were changed to emphasize awareness of natural conversation.

Procedure and Participants

In June 2017, the services of an internet research company were engaged, for the recruitment of predicted ICBT application users, while being mindful to find almost equal numbers of company employees, university students, and housewives. The selection process is outlined in Figure 1.

A total of 10,963 individuals were sent recruitment notices explaining, "SABORI is a self-monitoring service based on CBT, which is recognized for its high efficacy in mental health care," along with materials outlining the application and were asked to choose between "I would like to use it" and "I would not like to use it." Of these, 2668 individuals remained after removing the 8295 individuals who did not wish to use the application. Following this, explanatory sessions were offered over two days to explain the rules of use during the current month-long research study, namely the condition of using the application more than once every two days, totaling over 15 days of use. After removing 2109 individuals who were unable to participate, the remaining 559 individuals formed the experimental sample for the current research. After the one-month usage period, of the 427 responses from the poststudy questionnaire, it was revealed that 191 participants in the experimental group complied with the rule stipulating more than 15 days of use (46 male office workers, 34 female office workers, 13 male university students, 13 female university students, and 85 housewives; mean age 38.04 [SD 10.75] years). A total of 236 individuals used the application for less than 15 days. In addition, the 2109 individuals who were unable to participate in the one-month usage period were designated as the control group, and responses were closed when almost equivalent numbers of poststudy questionnaire responses were received, forming a control group of 263 individuals (51 male office workers, 53 female office workers, 26 male university students, 40 female university students, and 93 housewives; mean age 38.05 [SD 13.45] years). Moreover, members of the treatment group were paid the equivalent of US $4 worth of points from the internet research company for their participation. As a result, in response to the 6067 cases of user input accumulated during the study, the system's response success rate significantly improved to 92.93%.

Figure 1. Flow chart of participant selection.



Measures

The following measures were used to determine the mental health of the participants. Furthermore, a total of 48 scale items unrelated to the aims of the study, including items not mentioned here, were removed.

World Health Organization-Five Well-Being Index (WHO-5-J)

For this study, we utilized the Japanese version of the World Health Organization-Five Well-Being Index (WHO-5-J), which was developed by the World Health Organization; released by Psychiatric Research Unit, Psychiatric Center North Zealand [26]; and translated by Awata et al [27]. According to WHO-5's systematic reviews, it can be used as an outcome measure that balanced the wanted and unwanted effects of treatments; moreover, the index has been successfully applied across a wide range of study fields [28]. It is a one-factor scale which measures positive mental health, which is related to physical aspects that uses a 5-item, 6-point Likert scale. The reliability and validity of the Japanese version was confirmed by Awata et al [27]. It was employed as an evaluative index that utilized total scores, referring to participants' mental health over the past two weeks.

Kessler 10 (K10), Japanese Version

To measure the negative mental health, Kessler et al [29] created a one-factor scale related to physical aspects that used ratings of 10 items on a 5-point Likert scale. K10 has often been employed as an appropriate scale for measuring psychological distress in the Japanese population [30,31]. The reliability and validity of the Japanese version was confirmed by Furukawa et al [32]. It was employed as an effect index that utilized total scores of the past two weeks.

Behavioral Activation for Depression Scale (BADS), Japanese Version

A scale developed by Kanter et al [33] was used, which measured four factors: "Activation" (BADS-AC), "Avoidance/Rumination" (BADS-AR), "Work/School Impairment" (BADS-WS), and "Social Impairment" (BADS-SI) for over 25 items that were rated on a 7-point Likert scale. The scale is designed to evaluate functional impairment, avoidance, and activation in behavioral activation. BADS had a high validity in both a nondepressed sample [33] and in samples that had elevated depressive symptoms [34]. The Japanese version's reliability and validity was confirmed by Takagaki et al [35]. This version was employed as an effect index of activation (BADS-AC) for over 7 items, as well as avoidance/rumination (BADS-AR) over 8 items, for which total scores were utilized for each subscale that referred to the past two weeks. To support behavioral activation based on daily mood and physical condition, we improved the SABORI application and hypothesized that conversation-style free dialog promoted behavioral activation and reduced avoidance.

Results

Effect Index Descriptive Statistics

Table 1 details descriptive statistics for both groups' pre- and posttreatment scores.

Results of a Two-Factor Analysis of Variance (Group × Time)

A two-factor mixed model analysis of variance (ANOVA) was conducted to investigate differences for each effect index, group (treatment and control), and time (pre- and posttreatment) according to each scale (Table 2 and Figures 2-5). Moreover, error bars indicate a 95% confidence interval for the mean.

The two-way ANOVA (group and time independent variables) results indicated that SABORI had an effect on two variables: positive and negative mental health. The WHO-5 results indicated a significant main effect and interaction for time, $F_{1,452}$=5.79, P=.02; $F_{1,452}$=5.47, P=.02, respectively. After investigating the simple main effect of time for each level of the groups, a significant simple main effect was found for the treatment group, $F_{1,452}$=9.71, P=.002). On the K10, there was a significant main and interaction effects for time, $F_{1,452}$=5.43, P=.02; $F_{1,452}$=8.11, P=.005, respectively. Again, after investigating the simple main effect of time for each level of the groups, a significant simple main effect was found for the treatment group, $F_{1,452}$=11.57, P=.001. Importantly, the two-way ANOVA (independent variables: group × time) found a significant trend for behavioral activation, suggesting the potential for a certain degree of effectiveness. The BADS-AC results indicated that there was a significant main effect for time, $F_{1,452}$=2.75, P=.098. Investigation of the simple main effect of time for group each level indicated a significant simple main effect trend in the treatment group, $F_{1,452}$=3.53, P=.06. Neither a significant main nor interaction effects were found on the BADS-AR.

Table 1. Mean (SD) and Cronbach alpha of outcome measures. BADS-AC: Behavioral Activation for Depression Scale—Activation. BADS-AR: Behavioral Activation for Depression Scale—Avoidance/Rumination. K10: Kessler 10. WHO-5-J: WHO-Five Well-Being Index.

Measures	Control		Intervention	
	Mean (SD)	Cronbach alpha	Mean (SD)	Cronbach alpha
WHO-5-J				
Pretest	15.64 (5.53)	.91	15.03 (5.26)	.92
Posttest	15.65 (5.35)	.91	16.12 (4.71)	.89
K10				
Pretest	23.76 (9.97)	.95	23.58 (9.56)	.94
Posttest	23.97 (9.89)	.95	21.56 (8.26)	.93
BADS-AC				
Pretest	15.67 (8.27)	.89	16.09 (8.36)	.88
Posttest	15.84 (8.66)	.84	17.19 (7.90)	.84
BADS-AR				
Pretest	17.71 (8.75)	.91	18.51 (8.79)	.89
Posttest	18.35 (9.36)	.88	17.84 (8.03)	.82

Table 2. Effect of interaction between time and group on outcome variables. BADS-AC: Behavioral Activation for Depression Scale—Activation. BADS-AR: Behavioral Activation for Depression Scale—Avoidance/Rumination. K10: Kessler 10. WHO-5-J: WHO-Five Well-Being Index.

Measure	Main effect		Interaction effect	Simple main effect (partial η^2)	
	Time	Group	Time × group	Control	Intervention
WHO-5-J	5.79	0.02	5.47	0.00	9.71 (.02)
K10	5.43	2.52	8.11	0.16	11.57 (.03)
BADS-AC	2.75	1.45	1.66	0.12	3.53 (.01)
BADS-AR	0.00	0.40	2.39	1.35	1.08

[a]WHO-5-J: WHO-Five Well-Being Index.

[b]K10: Kessler 10.

[c]BADS-AC: Behavioral Activation for Depression Scale—Activation.

[d]BADS-AR: Behavioral Activation for Depression Scale—Avoidance/Rumination.

Figure 2. Change in the World Health Organization-Five Well-Being Index.

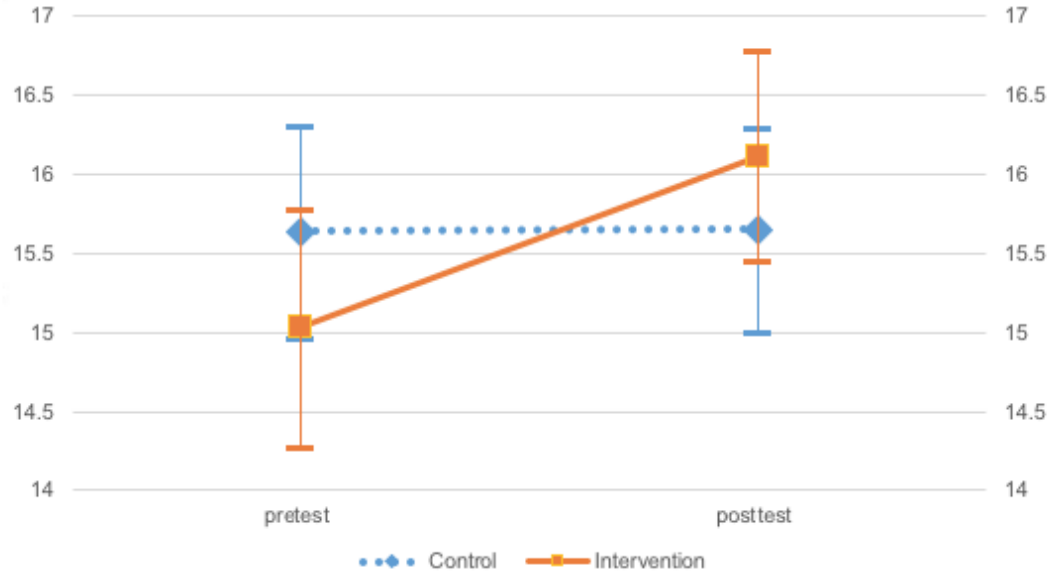


Figure 3. Change in Kessler 10.

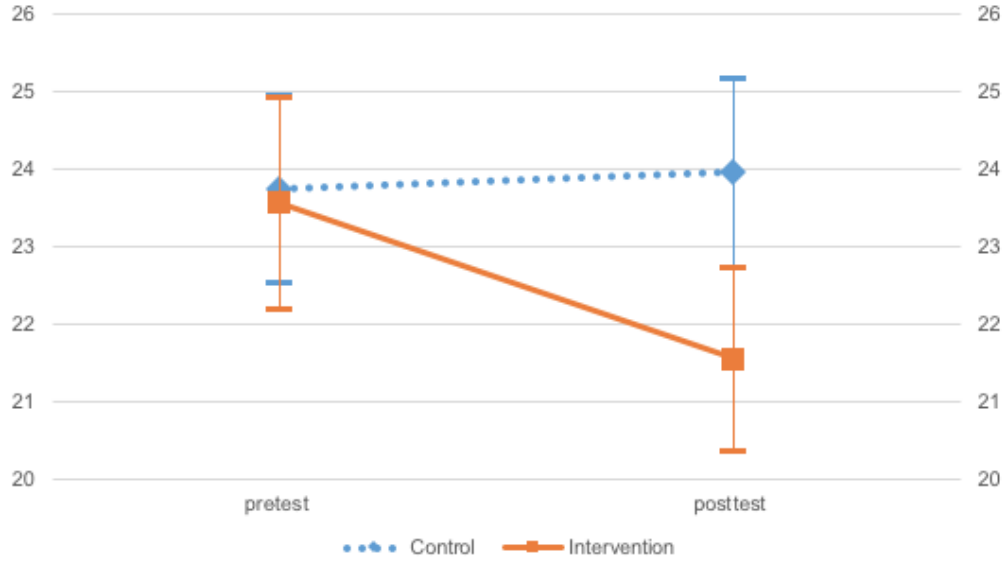


Figure 4. Change in the Behavioral Activation for Depression Scale—Activation.

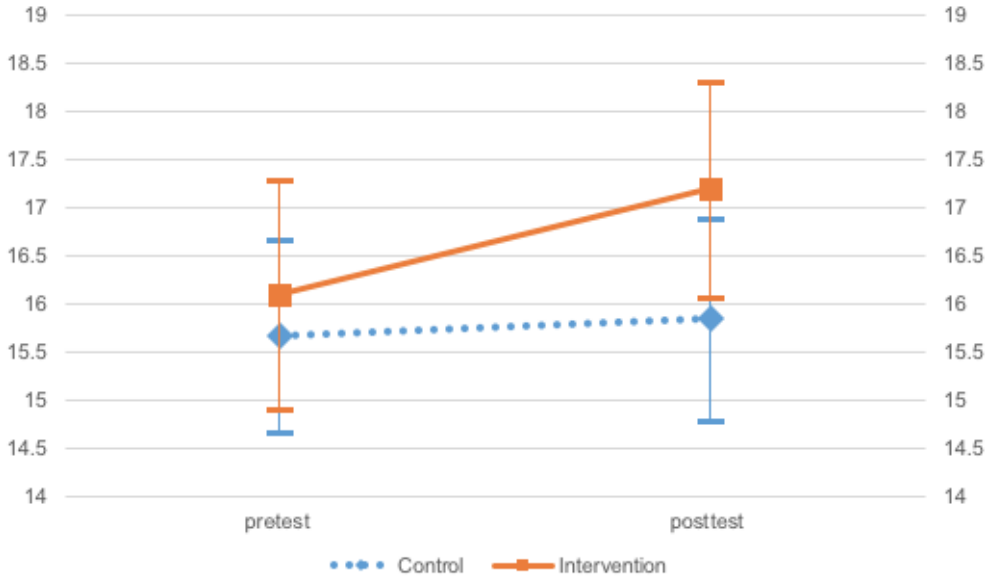
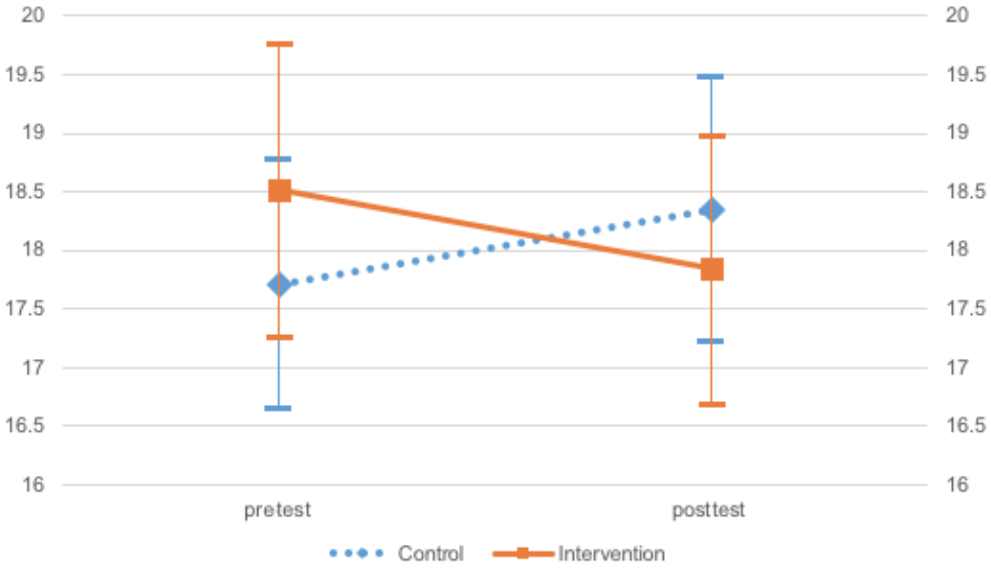


Figure 5. Change in the Behavioral Activation for Depression Scale—Avoidance/Rumination.



Discussion

Positive and Negative Effects on Mental Health

Accordingly, this can be seen as evidence suggesting the preliminary efficacy of the agent-based conversation-style mental health self-care application, SABORI, across a wide range of areas, due to the indicated effects on both positive and negative mental health. Furthermore, research into the preliminary efficacy of the version of SABORI without the conversation-style dialog feature [25] found that it improved negative mental health for depression only, whereas the current research indicated that the new version of SABORI application with an ECA is effective beyond just depression, acting across a wide range of areas, including preventive mental health measures. The addition of the agent-based dialog feature potentially affected the strengthening of the therapeutic alliance between the system and the user. But there is a need for further detailed research into the factors underlying the effect, as well as the component factors of the therapeutic alliance when utilizing agent. Agent, especially the ECA, is a technology that has the potential to vastly change traditional mental health support measures, through its use as an alternative to face-to-face interviews. The study's results are of major significance to future development because they suggest it may be possible to add an ECA to unguided ICBT, which was traditionally criticized on grounds of low effectiveness and a paucity of evidence. Nevertheless, in light of the still inadequate efficacy level, we must continue to seek ways to increase efficacy by linking to educational psychology content and delivering optimal individualization of the dialog-based behavior suggestion feature, along with further improvement of knowledge-base of ECA with the data collected in this study. Furthermore, the ECA in the SABORI focused on texts and pictures, which can also have audio and visual features. For example, connecting SABORI with a smart speaker could improve usability and enhance outcomes.

Efficacy Regarding Behavioral Activation, Avoidance, and Rumination

One possible reason for the weakness of the behavioral activation effect was the one-month usage time frame of the current research. Although the new version of SABORI offers behavioral suggestions through the dialog feature based on current circumstances, these function as a catalyst at most, and the intent is that, with continued usage, the user will become able to undertake behavioral activation. However, it is conceivable that the one-month time frame is too short for voluntary behavioral activation to occur, so the effect never extends further than carrying out the behavioral suggestions. Future research will need to investigate how behavioral activation changes over a longer period of time, based on monthly use.

In addition, no effect was found for behavioral activation related factors of avoidance and rumination. There are two possible reasons for this. First, similar to behavioral activation, the one-month time frame may have been too short to elicit a behavioral change, revealing the necessity for a further longer-term study. Second, it is conceivable that the use of healthy participants may have contributed to this outcome. Since the newly developed SABORI is intended for use as a mental health preventative measure, it has no particular audience limitations. Therefore, it is possible that any effect could not be adequately measured in a healthy cohort, given that behavioral activation treatment is primarily intended as a treatment for depression, and the BADS used in the current research as an effect index also contains many items relating to depression. Consequently, it is hoped that the efficacy of the ICBT application with an ECA will be even more comprehensively examined by conducting research utilizing effect indices that are more suitable in terms of mental health preventative measures intended for a healthy cohort.

Limitations

This study was a pilot trial for assessing feasibility and acceptability, and the results suggested that this version of SABORI is promising for preventive mental health; however, there are certain limitations. First, it is impossible to conclude that the ECA enhances therapeutic alliance; thus, further research is required to evaluate the therapeutic alliance between users and agents in the SABORI. The second limitation is related to the research design. This study does not show the exact influence on the result related to SABORI's efficacy. We compared only the control group and the experimental group in which users were instructed to use the SABORI. Because the SABORI consisted of ICBT and ECA, only the experimental group was required to clarify the ECA's efficacy. Furthermore, because the current study involved only a nonrandomized comparison, it is undeniably inferior in terms of the evidence quality compared to evidence from randomized comparison research. Consequently, using the results of this study, it was necessary to pursue a higher level of evidence by conducting a randomized comparison study and adding different participant groups. Third, the study did not deal with a clinical population; therefore, it is impossible to conclude that the SABORI is useful for such populations. Thus, further research focused on a clinical population is necessary.

Conclusions

The current research provides evidence regarding feasibility and acceptability in mental health care delivered by the SABORI, a self-care application utilizing AI technology, especially agent technology. To date, unguided ICBT has failed to garner attention, being considered ineffective with high dropout rates [10,11]. However, current research has revealed that it is possible to use ECA to compensate for clinical failures of ICBT such as the impoverished therapeutic alliance. If the technological development of conversational agent continues to progress, it could be possible to form a therapeutic alliance with agent to rival that found in face-to-face therapy. The utilization of agent technology is anticipated to vastly change the traditional, mainstream delivery of mental health services and could be an innovation that holds the secret to filling the service gap [3]. Moreover, it is expected to contribute greatly to clinical practice and be increasingly utilized in the area of mental health care.

Acknowledgments

We would like to thank Yuta Tonoike and Daiki Morita of Mind Isle Co Ltd for their cooperation in this research study. This research was supported by the Graduate Program for Social ICT Global Creative Leaders of The University of Tokyo and by the Japan Society for the Promotion of Science.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the monitoring questions.

[PDF File (Adobe PDF File), 49KB - mental_v5i3e10454_app1.pdf]

Multimedia Appendix 2

Screenshot of the behavioral suggestion dialog.

[PDF File (Adobe PDF File), 81KB - mental_v5i3e10454_app2.pdf]

References

1. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 8;196(4286):129-136. [Medline: 847460]
2. Westbrook D, Kennerley H, Kirk J. An introduction to cognitive behaviour therapy: skills and applications, second edition. London: SAGE Publications; 2011.
3. Stefl ME, Prosperi DC. Barriers to mental health service utilization. Community Ment Health J 1985;21(3):167-178. [Medline: 4075773]
4. Bardram J, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Vedel KL. Designing mobile health technology for bipolar disorder: a field trial of the Monarca system. New York: ACM; 2013 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13); April 27 - May 02, 2013; Paris p. 2627-2636. [doi: 10.1145/2470654.2481364]
5. Bardram J, Frost M, Szántó K, Marcu G. The MONARCA self-assessment system: a persuasive personal monitoring system for bipolar patients. 2012 Presented at: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12); January 28 - 30; Miami p. 21-30. [doi: 10.1145/2110363.2110370]
6. Coyle D, McGlade N, Doherty G, O'Reilly G. Exploratory evaluations of a computer game supporting cognitive behavioural therapy for adolescents. 2011 Presented at: the SIGCHI Conference on Human Factors in Computing Systems (CHI '11); May 07 - 12; Vancouver p. 2937-2946. [doi: 10.1145/1978942.1979378]
7. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. 2012 Presented at: the SIGCHI Conference on Human Factors in Computing Systems (CHI '12); May 05 - 10; Austin p. 1421-1430. [doi: 10.1145/2207676.2208602]
8. Matthews M, Doherty G. In the mood: engaging teenagers in psychotherapy using mobile phones. 2011 Presented at: the SIGCHI Conference on Human Factors in Computing Systems (CHI '11); May 07 - 12; Vancouver p. 2947-2956. [doi: 10.1145/1978942.1979379]
9. Rennick-Egglestone S, Knowles S, Toms G, Bee P, Lovell K, Bower P. Health Technologies 'In the Wild': Experiences of Engagement with Computerised CBT. 2016 Presented at: the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16); May 07 - 12; San Jose p. 2124-2135. [doi: 10.1145/2858036.2858128]
10. Andersson G. The Internet and CBT: A Clinical Guide. Boca Raton: CRC Press; 2014.
11. Andersson G. Internet-Delivered Psychological Treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
12. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
13. Elvins R, Green J. The conceptualization and measurement of therapeutic alliance: an empirical review. Clin Psychol Rev 2008 Oct;28(7):1167-1187. [doi: 10.1016/j.cpr.2008.04.002] [Medline: 18538907]
14. Barazzone N, Cavanagh K, Richards DA. Computerized cognitive behavioural therapy and the therapeutic alliance: a qualitative enquiry. Br J Clin Psychol 2012 Nov;51(4):396-417. [doi: 10.1111/j.2044-8260.2012.02035.x] [Medline: 23078210]
15. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
16. Beun RJ, de Vos E, Witteman C. Embodied Conversational Agents: Effects on Memory Performance and Anthropomorphisation. In: Rist T, Aylett R, Ballin D, editors. Intelligent Virtual Agents. IVA 2003. Lecture Notes in Computer Science. Berlin: Springer; 2003:315-319.
17. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
18. Martínez-Miranda J. Embodied Conversational Agents for the Detection and Prevention of Suicidal Behaviour: Current Applications and Open Challenges. J Med Syst 2017 Sep;41(9):135. [doi: 10.1007/s10916-017-0784-6] [Medline: 28755270]
19. Sebastian J, Richards D. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. Computers in Human Behavior 2017 Aug;73:479-488. [doi: 10.1016/j.chb.2017.03.071]
20. Yokotani K, Takagi G, Wakashima K. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. Computers in Human Behavior 2018 Aug;85:135-145. [doi: 10.1016/j.chb.2018.03.045]
21. Woebot Labs Inc. 2018 URL: https://woebot.io [accessed 2018-05-16]
22. X2AI Inc. 2018 URL: http://x2ai.com/ [accessed 2018-05-16]
23. Hudlicka E. Virtual Affective Agents and Therapeutic Games. In: Luxton DD, editor. Artificial Intelligence in Behavioral and Mental Health Care. Cambridge: Academic Press; 2015:81-116.

24. Blümle A, Schandelmaier S, Oeller P, Kasenda B, Briel M, von EE, DISCO study group. Premature Discontinuation of Prospective Clinical Studies Approved by a Research Ethics Committee - A Comparison of Randomised and Non-Randomised Studies. PLoS One 2016;11(10):e0165605 [FREE Full text] [doi: 10.1371/journal.pone.0165605] [Medline: 27792749]
25. Hirano M, Ogura K, Kitahara M, Sakamoto D, Shimoyama H. Designing behavioral self-regulation application for preventive personal mental healthcare. Health Psychol Open 2017 Jan;4(1):2055102917707185 [FREE Full text] [doi: 10.1177/2055102917707185] [Medline: 28567301]
26. Psychiatric RU, Psychiatric CNZ. WHO5_Japanese. 2017 URL: https://www.psykiatri-regionh.dk/who-5/Documents/WHO5_Japanese.pdf [accessed 2018-03-20] [WebCite Cache ID 6y3NnWZkX]
27. Awata S, Bech P, Yoshida S, Hirai M, Suzuki S, Yamashita M, et al. Reliability and validity of the Japanese version of the World Health Organization-Five Well-Being Index in the context of detecting depression in diabetic patients. Psychiatry Clin Neurosci 2007 Feb;61(1):112-119 [FREE Full text] [doi: 10.1111/j.1440-1819.2007.01619.x] [Medline: 17239048]
28. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015 Mar;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
29. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [Medline: 12214795]
30. Konishi M, Tachibana Y, Tang J, Takehara K, Kubo T, Hashimoto K, et al. A Comparison of Self-Rated and Female Partner-Rated Scales in the Assessment of Paternal Prenatal Depression. Community Ment Health J 2016 Dec;52(8):983-988. [doi: 10.1007/s10597-015-9931-z] [Medline: 26308837]
31. Honda A, Abe Y, Date Y, Honda S. The Impact of Multiple Roles on Psychological Distress among Japanese Workers. Saf Health Work 2015 Jun;6(2):114-119 [FREE Full text] [doi: 10.1016/j.shaw.2014.12.004] [Medline: 26106510]
32. Furukawa TA, Kawakami N, Saitoh M, Ono Y, Nakane Y, Nakamura Y, et al. The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. Int J Methods Psychiatr Res 2008;17(3):152-158. [doi: 10.1002/mpr.257] [Medline: 18763695]
33. Kanter J, Mulick P, Busch A, Berlin K, Martell C. The Behavioral Activation for Depression Scale (BADS): Psychometric Properties and Factor Structure. J Psychopathol Behav Assess 2006 Oct 25;29(3):191-202. [doi: 10.1007/s10862-006-9038-5]
34. Kanter JW, Rusch LC, Busch AM, Sedivy SK. Validation of the Behavioral Activation for Depression Scale (BADS) in a Community Sample with Elevated Depressive Symptoms. J Psychopathol Behav Assess 2008 Jul 24;31(1):36-42. [doi: 10.1007/s10862-008-9088-y]
35. Takagaki K, Okajima I, Kunisato Y, Nakajima S, Kanai Y, Ishikawa S, et al. Development and validation of the Japanese version of the Behavioral Activation for Depression Scale (BADS). Archives of psychiatric diagnostics and clinical evaluation 2013;6:76-85 [FREE Full text]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
BA: behavioral activation
BADS: Behavioral Activation for Depression Scale
BADS-AC: Behavioral Activation for Depression Scale—Activation
BADS-AR: Behavioral Activation for Depression Scale—Avoidance/Rumination
BADS-WS: Behavioral Activation for Depression Scale—Work/School Impairment
BADS-SI: Behavioral Activation for Depression Scale—Social Impairment
CBT: cognitive behavioral therapy
ECA: embodied conversational agent
ICBT: internet-based cognitive behavioral therapy
K10: Kessler 10
NPS: nonrandomized prospective study
WHO-5-J: World Health Organization-Five Well-Being Index

Edited by J Torous; submitted 20.03.18; peer-reviewed by V Bhat, S D'Alfonso; comments to author 01.05.18; revised version received 25.06.18; accepted 11.07.18; published 31.07.18.

Please cite as:
Suganuma S, Sakamoto D, Shimoyama H
An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial
JMIR Ment Health 2018;5(3):e10454
URL: http://mental.jmir.org/2018/3/e10454/
doi:10.2196/10454
PMID:30064969

Original Paper

Emotion Recognition Using Smart Watch Sensor Data: Mixed-Design Study

Juan Carlos Quiroz[1,2], PhD; Elena Geangu[3], PhD; Min Hooi Yong[4], PhD

[1]Department of Computer Science and Engineering, University of Nevada, Reno, NV, United States
[2]Australian Institute of Health Innovation, Macquarie University, NSW, Australia
[3]Department of Psychology, University of York, York, United Kingdom
[4]Department of Psychology, Sunway University, Bandar Sunway, Malaysia

Corresponding Author:
Min Hooi Yong, PhD
Department of Psychology
Sunway University
5 Jalan Universiti
Bandar Sunway, 47500
Malaysia
Phone: 60 374918622
Fax: 60 356358633
Email: mhyong@sunway.edu.my

Abstract

Background: Research in psychology has shown that the way a person walks reflects that person's current mood (or emotional state). Recent studies have used mobile phones to detect emotional states from movement data.

Objective: The objective of our study was to investigate the use of movement sensor data from a smart watch to infer an individual's emotional state. We present our findings of a user study with 50 participants.

Methods: The experimental design is a mixed-design study: within-subjects (emotions: happy, sad, and neutral) and between-subjects (stimulus type: audiovisual "movie clips" and audio "music clips"). Each participant experienced both emotions in a single stimulus type. All participants walked 250 m while wearing a smart watch on one wrist and a heart rate monitor strap on the chest. They also had to answer a short questionnaire (20 items; Positive Affect and Negative Affect Schedule, PANAS) before and after experiencing each emotion. The data obtained from the heart rate monitor served as supplementary information to our data. We performed time series analysis on data from the smart watch and a *t* test on questionnaire items to measure the change in emotional state. Heart rate data was analyzed using one-way analysis of variance. We extracted features from the time series using sliding windows and used features to train and validate classifiers that determined an individual's emotion.

Results: Overall, 50 young adults participated in our study; of them, 49 were included for the affective PANAS questionnaire and 44 for the feature extraction and building of personal models. Participants reported feeling less negative affect after watching sad videos or after listening to sad music, $P<.006$. For the task of emotion recognition using classifiers, our results showed that personal models outperformed personal baselines and achieved median accuracies higher than 78% for all conditions of the design study for binary classification of happiness versus sadness.

Conclusions: Our findings show that we are able to detect changes in the emotional state as well as in behavioral responses with data obtained from the smartwatch. Together with high accuracies achieved across all users for classification of happy versus sad emotional states, this is further evidence for the hypothesis that movement sensor data can be used for emotion recognition.

(JMIR Ment Health 2018;5(3):e10153) doi:10.2196/10153

KEYWORDS

emotion recognition; accelerometer; supervised learning; psychology

Introduction

Our emotional state is often expressed in a variety of means, such as face, voice, body posture, and walking gait [1,2]. Many studies are conducted in strict laboratory settings, which may impede the overall ecological validity of the findings. Having a strong ecological validity is important because emotional expression or display in any modality is not entirely dependent on conscious action or function. Instead, emotional expressions are essentially a response to a particular affective stimulus or experience, and this response might be reduced in a laboratory as a result of social desirability.

Speech, video, and physiological data have been analyzed to determine the emotional state of a person [3,4], but these analyses usually rely on recordings obtained in laboratory environments with limited ecological validity. To formulate theoretical models of emotions and affective health that take into account the richness of everyday life, we need to measure affective states unobtrusively. Mobile phones include sensors, such as accelerometers, that have the potential to be sensitive to changes in people's affective states and thus could provide rich and accessible information in this respect; for example, we know that the way we walk reflects whether we feel happy or sad [2]. This paper analyzes movement sensor data recorded via a smart watch in relation to changes in emotions.

Prior work on emotion detection from mobile phone data includes analysis of typing behavior on a mobile phone [5,6] and mobile phone usage [7,8]. The EmotionSense system performed emotion detection directly on mobile phones via analysis of speech with additional sensors collecting information about the user and the environment [9]. However, there are some indications that movement sensor data collected by mobile phones could be a viable solution for inferring emotion, as opposed to inferring movement or physical activities. Cui et al attempted to record participants' movements with mobile phones strapped to their ankles and wrists, thus impairing ecological validity [10]. Happiness and anger were elicited with video stimuli, and emotional state classifiers were trained with accelerometer data [10]. Zhang et al also focused on happiness and anger, but they recorded movement data with smart bracelets [11]. Accuracies in detecting these emotions ranged from 60% to 91.3% across all subjects using 10-fold cross-validation [11].

These cases have motivated further research on tracking and analysis of sensor data from mobile phones and wearables with the goal of monitoring and intervening for patients suffering from mental illnesses or substance abuse [12,13]. Further validation is needed for the hypothesis that movement sensor data can be used to recognize emotional states. Movement data are of particular interest because accelerometers and gyroscopes are standard sensors in mobile phones, wearables, and fitness trackers. Movement data collection is unobtrusive, and it requires no user input [14], which gives us reliable data in the real world without the possibility of having social desirability responses.

Toward the end, we make the following contributions. First, we conducted a mixed-design study, as seen in Figure 1, with 50 participants to test two types of stimuli, audiovisual and audio, for eliciting emotional responses from participants. Participants wore a smart watch on the wrist and a heart rate strap on the chest. The heart rate strap was included to supplement data collected from Positive Affect and Negative Affect Schedule (PANAS) scores [15]. After or while watching emotional stimuli, participants walked, and the process was repeated three times, for each of the following emotions: happy, neutral, and sad. We extracted features from sensor data and built classifiers (personal models) that recognized the emotional state of the user. Our results show that personal models outperformed personal baselines and achieved median accuracies higher than 78% for all conditions of the design study for binary classification of happiness versus sadness. This paper is an extended version of preliminary findings published [16].

Figure 1. Mixed-design study with three conditions. The three conditions were used to determine the stimulus that would better induce the target emotional states on participants. PANAS: Positive Affect and Negative Affect Schedule.

Methods

Participants

In total, 50 young adults participated in this study (43 females; mean age 23.18 [SD 4.87] years). All participants were recruited in a university campus (North-West UK) via announcements on notice boards and by word of mouth. Each participant was given £7 for participation. None of the participants reported any vision or hearing difficulties and could walk unassisted.

Ethics Approval

We obtained ethical approval from Sunway University Ethics Board (SUREC 2016/05) and had it validated by Lancaster University to conduct both validation and the actual main study experiment.

Materials

The study included the following two types of stimuli: a) audiovisual and b) audio.

Audiovisual

Audiovisual clips were selected from commercial movies with the potential of being perceived as having emotional meaning (ie, sadness and happiness) and able to elicit emotional responses. Commercial movies were selected from Gross and Levenson [17], Bartollini [18], Schaefer et al [19], and from 5 young adults (4 females; mean age 21.50 years). Another 5 participants (3 females; mean age 22.80 [SD 1.30] years) were asked to identify each of these clips in terms of the emotion they felt while watching, and the intensity of the emotion they felt using a 0-to-10 Likert scale (0: hardly; 10: very much likely). They were also asked if they had watched that movie before. On average, only one participant had seen that movie before. Participants reported that they felt the emotion intended for all clips (100% accuracy) and the intensity experienced ranged between 5.0 to 6.5 for happy and sad clips, respectively. Table 1 includes the movie clips used in our study.

Audio

For audio stimuli, pieces of classical music known to elicit happy, sad, and emotionally neutral states were chosen [20]. Table 2 includes selected clips.

Procedure

All participants were presented with happy, sad, and neutral stimuli. A third of the participants (n=18) were presented with audiovisual stimuli (ie, videos), whereas the other participants (n=32) were presented with audio stimuli (ie, classical music). Half the participants (n=16), who were assigned to audio stimuli, listened to them prior to walking, whereas the other half (n=16) listened to stimuli while they were walking. Eighteen participants (n=18) who were assigned to watch emotional videos watched them prior to walking. Assignment to each condition was random. To counter possible order effects, half the participants were presented with sad stimuli first, whereas the other half were presented with happy stimuli first. Each participant was tested individually, and the task took approximately 20 minutes to complete. All data was collected between 17:00 and 19:00 h to account for peak foot traffic.

Table 1. Movies used to induce happy and sad emotions.

Emotion and movie	Scene
Happy	
10 Things I Hate About You (1999)	Patrick serenades Katarina in stadium
When Harry Met Sally (1989)	Discussion of orgasms in cafe
There's Something about Mary (1998)	Mary hair gel scene
Monty Python (1975)	Black Knight fights King Arthur
Modern Times (1936)	Factory worker in assembly line
Love Actually (2003)	Arrival halls scene in Heathrow airport
Wall-E (2008)	EVA kisses Wall-E
Benny & Joon (1993)	Sam roll dance in diner
Sad	
Interstellar (2014)	Cooper watches video messages sent by his children
Click (2006)	Michael rewinds his past to recall not saying goodbye to his father
Hachi (2009)	Hachiko waits at the train station
Shawshank Redemption (1994)	Death of Brooks
Saving Private Ryan (1998)	Mother is informed of the deaths of all of Private Ryan's brothers
Marley & Me (2008)	Marley is euthanized in the veterinarian clinic
The Champ (1979)	Boy cries at father's death
My Girl (1991)	Thomas's funeral

Table 2. Musical pieces used to induce happy and sad emotions and neutral ones.

Emotion and piece	Composer
Happy	
Carmen: "Chanson du Toreador"	Bizet
"Allegro"—*A Little Night Music*	Mozart
"Rondo Allegro"—*A Little Night Music*	Mozart
"Blue Danube"	Strauss
"Radetzky March"	Strauss
Sad	
Adagio in Sol Minor	Albinoni
"Kol Nidrei"	Bruch
"Solveig's song"—*Peer Gynt*	Grieg
Concierto de Aranjuez	Rodrigo
Suite for violin & orchestra in A minor	Sinding
Neutral	
"L'oiseau prophete"	Schumann
"Au Clair de lune"	Beethoven
"Clair de lune"	Debussy
Symphony no. 2 in C minor	Mahler
La Traviata	Verdi
Pictures at an Exhibition	Mussorgsky
"Water Music Suite: 5. Passepied"	Handel
"Violin Romance no. 2 in F major"	Beethoven
"Water Music"—minuet	Handel
The Planets —"Venus"	Holst

The three conditions of the mixed-design study are presented in Figure 1 and are as follows: Condition 1—watching the movie clip prior to walking; Condition 2—listening to the music prior to walking; and Condition 3—listening to the music while walking.

Each participant was first greeted by the experimenter at one end of the corridor and was helped to put on various items. First, the participant had the heart rate sensor (Polar H7) strapped snugly around the chest. The corresponding watch (Polar M400) was strapped onto the experimenter's wrist. The watch was set to the "other indoor" sport profile. Second, the participant strapped a smart watch (Samsung Gear 2) on the left wrist. Participants wore sensors for the entire duration of the experiment. The smart watch included a triaxial accelerometer and a triaxial gyroscope. The sampling rate of the smart watch is advertised as 25 Hz, but our results show that the actual sampling rate on average was 23.8 Hz. For the smart watch, we developed a Tizen app that recorded accelerometer and gyroscope sensor data.

Participants rated their current mood state using PANAS [21] on a 7-inch tablet. PANAS contains 10 adjectives for positive (eg, joy) and 10 adjectives for negative feelings (eg, anxiety). Scores can range from 10 to 50 with higher scores representing higher levels of affect. The heart rate sensor was used in the study to supplement data collected from PANAS scores [15].

For Conditions 1 and 2, in which the stimulus presentation occurred before walking, participants wore a pair of headphones to listen or watch the assigned stimuli (eg, sad music or happy movie) while at the start of a walking route. At the end of the stimulus, the participant walked to the end of the route and back to the starting point. Participants were reminded not to make any stops in between. The route was represented by a 250 m S-shaped corridor located on the ground floor of a university building. The experimenter discreetly followed the participant at a 125 m distance to observe the behavior and to ensure that heart rate monitoring was captured by the watch. Upon return, participants rated their mood using the same PANAS scales. Because of the initial mood induction, we always had a neutral condition between happy and sad conditions to allow return to the baseline calm state. For all participants, the neutral stimulus was classical music for the audio type or a movie with classical music playing in the background and depicting an everyday scene. The same procedure above, rating their initial mood using PANAS, watching or listening to a stimulus, walking along the corridor and back, and rating their mood, was applied to the neutral and second emotion.

In Condition 3, which included listening while walking, the procedure was similar as above, except that the participant was listening to the assigned music while walking, and participants reported PANAS scores after walking.

Feature Extraction

During the experiment, the experimenter recorded the time each participant started and stopped walking. These times were used to identify sensor data that corresponded to the actual walking time. We discarded sensor data when participants were briefed and when participants watched or listened to the stimulus prior to walking.

The walking times were labeled according to the corresponding emotional stimulus presented before walking; for example, if the participant viewed a movie clip known to induce happiness, all of the features extracted from the subsequent walking data were labeled as happy. These labels were used to train classifiers for the recognition of happiness versus sadness. We present classifier results for the two-class problem of detecting happy versus sad emotions and for the three-class problem of detecting happy versus sad versus neutral emotions.

We first filtered raw accelerometer data with a mean filter (window=3). Features were extracted from sliding windows with a size of one second (24 samples) with 50% overlap. Our feature extraction approach is similar to that used for activity recognition from mobile phone accelerometer data [22,23], that is, each window is treated as an independent sample (feature vector). We address neighborhood bias when building models from accelerometer sliding windows in the results section [24].

For each window of the triaxial accelerometer and triaxial gyroscope data, we extracted 17 features [23]: (1) mean, (2) SD, (3) maximum, (4) minimum, (5) energy, (6) kurtosis, (7) skewness, (8) root mean square, (9) root sum square, (10) sum, (11) sum of absolute values, (12) mean of absolute values, (13) range, (14) median, (15) upper quartile, (16) lower quartile, and (17) median absolute deviation. These 17 features were extracted from each of the 3 axes of the accelerometer data and each of the 3 axes of the gyroscope data, resulting in 102 features. We also calculated the angle between the signal mean (within a window) and the x-axis, y-axis, and z-axis (3 features); SD of signal magnitude (one feature); and the heart rate (one feature) for a total of 107 features for the feature vector of a window. Unless stated otherwise, we used all 107 features for classification. However, we do explore classification performance based on features corresponding to certain sensors: accelerometer, gyroscope, and heart rate; accelerometer and heart rate; and accelerometer.

We divided data by condition and built personal models with features extracted from each window [25]. In personal models, training and testing data come from a single user. In our case, we built 44 personal models (data from 6 participants were discarded because of missing data and other recording errors) with each model evaluated using stratified 10-fold cross-validation that was repeated 10 times. For each participant, we obtained a mean of 403.29 (55.62) samples labeled as happy, 403.67 (51.46) samples labeled as sad, and 402.93 (50.24) samples labeled as neutral. Of the 44 personal models built, 16 were from Condition 1 (watch movie and then walk), 14 were from Condition 2 (listen to music and then walk), and 14 were from Condition 3 (listen to music and then walk).

We compared random forest models, with 100 estimators and logistic regression, with L2 regularization and a baseline classifier that picked the majority class as the prediction. The python scikit-learn library was used for training and testing these classifiers. Because the number of samples labeled as happy versus sad for each participant was approximately the same, the baseline classifier predicted each window as happy versus sad with about a 50% probability (ie, all samples for user *i* were classified as happy, resulting in about 50% accuracy). For binary classification of happy versus sad, we use the accuracy, the F1 score, and the area under the receiver operating characteristic curve (ROC AUC) to assess classification performance. For multiclass classification of happy versus sad versus neutral, we use the accuracy and the F1 score.

Results

Ecological Validity Checks

When asked about their experience in using a smart gadget, most participants were familiar and comfortable with the smart watch but not with the Polar heart rate monitor. They did not notice anything unusual about the study that might have influenced their walking gait and behavioral response.

Behavioral Response to Stimuli (Positive Affect and Negative Affect Schedule Outcomes)

We analyzed PANAS responses for all conditions on the happy versus sad stimuli. One participant's data was excluded for being incomplete, thus leaving 49 for analyses (15 for Condition 1, 18 for Condition 2, and 16 for Condition 3). We first reviewed normality and found that data was normally distributed for Conditions 1 and 2 but not for Condition 3 (visual histograms were skewed and Shapiro-Wilk $P<.01$). See Multimedia Appendix 1 for PANAS scores for each emotion.

Condition 1: Watch Movie and Then Walk

Participants reported a reduced negative affect after watching a sad movie clip (mean 14.94 [SD 6.79]) compared with that before watching it (mean 19.00 [SD 7.20], $t_{16}=3.16$, $P=.006$). There was no significant difference for the positive affect for the sad movie ($t_{16}=.08$, $P=.94$) and for both affects with respect to the other two emotions (happiness and neutral), all P values were >.10.

Condition 2: Listen to Music and Then Walk

For sad music, participants reported an increased positive effect after the walk (mean 24.00 [SD 5.33]) compared with that before watching it (mean 20.31 [SD 5.79], $t_{15}=2.96$, $P=.01$) and reduced negative affect after (mean 11.69 [SD 3.34]) compared with that before watching (mean 13.63, [SD 5.12], $t_{15}=2.78$, $P=.014$). Participants reported reduced positive affect after listening to happy music (mean 26.38 [SD 6.96]) compared with that before watching it (mean 29.56 [SD 5.17], $t_{15}=2.62$, $P=.02$), but no significant difference for negative affect ($t_{15}=1.60$, $P=.13$). There

was no significant difference for neutral music for both affects, both *P* values were >.76.

Condition 3: Listen to Music While Walking

Participants reported an increased negative affect while walking and listening to happy music (mean 13.31 [SD 4.88]) compared with neutral (mean 15.00 [SD 5.44]) music (Z=2.64, *P*=.08). No other significant differences were observed, all *P* values were >.13.

Heart Rate

We planned to verify data obtained from PANAS and determine whether our participants experienced accelerated or decelerated heart rate as a result of emotional stimuli [15]. From the 50 participants, we had some data loss due to technical faults (n=9; 3 from Condition 1, 3 from Condition 2, and 3 from Condition 3), leaving us with data obtained from 41 participants. We first reviewed descriptive statistics and found that data was normally distributed. A one-way between-subjects analysis of variance was conducted to compare the effect of emotion (happy, sad, and neutral) on participants' heart rates. We did not find any significant effect of emotion on their heart rate for the 3 conditions ($F_{2,120}$=0.13, *P*=.88; see Table 3 for means and SD).

Emotion Recognition

Happy Versus Sad

Figure 2 illustrates 3 boxplot sets, one for each condition, showing distribution of accuracies for the personal model of each participant. For all 3 conditions, personal baselines have accuracies in the range 50%-54%. For all conditions, both random forest model and logistic regression outperformed the baseline with accuracies in the range 62%-99%. Condition 1 (movie) and Condition 3 (music while walking) resulted in the highest classification accuracies with median accuracies over 82%.

Table 3. Mean heart rate and SD in brackets for all 3 emotions.

Emotions	Mean (SD)
Happy	104.43 (14.55)
Sad	91.68 (16.31)
Neutral	105.77 (14.50)

Figure 2. Boxplot of classification accuracies for participants divided by conditions. Algorithms tested were baseline (pick majority), random forests, and logistic regression. Outliers are indicated by +. The highest classification accuracies were achieved with Condition 1 (movie) and Condition 3 (music while walking). For all conditions, the models achieved accuracies greater than 78% for over half the users. RF: random forest.

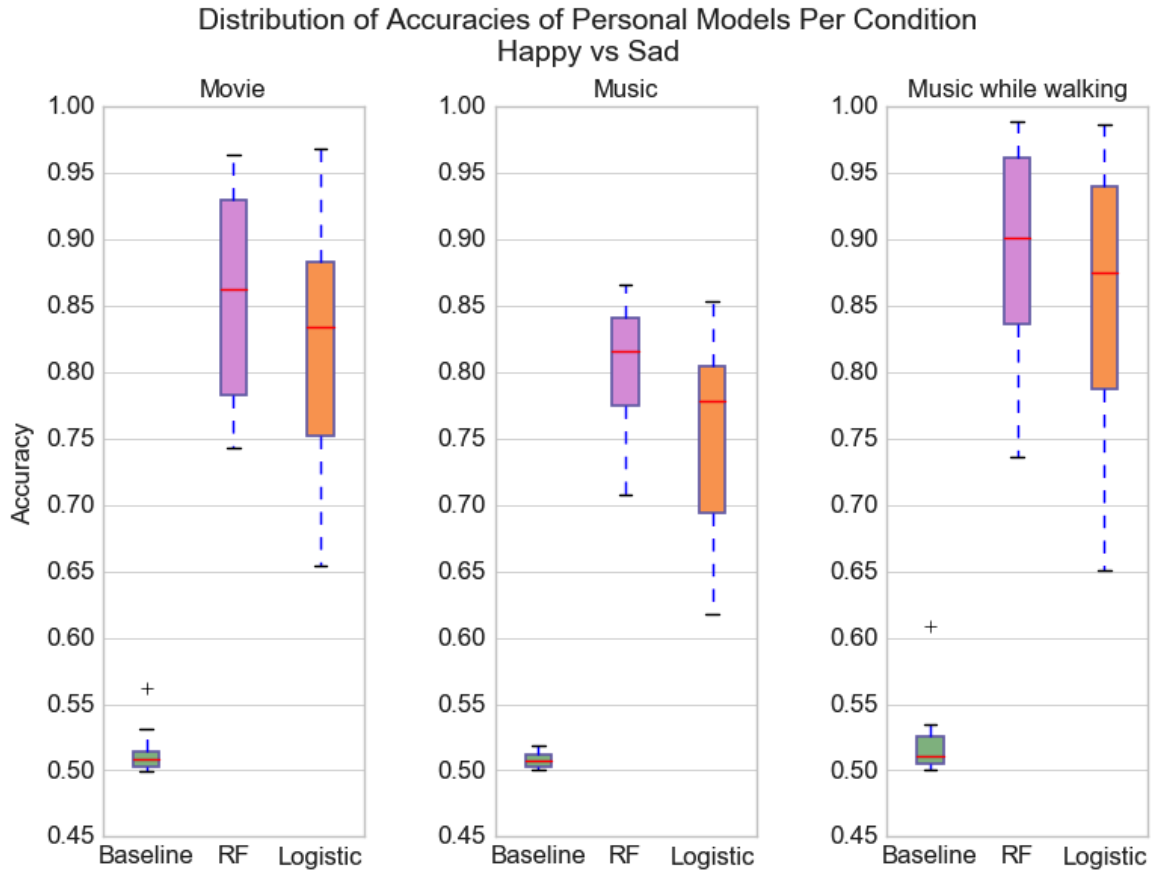


Table 4. Average user lift and average personal model accuracy per condition.

Features and model	AUC[a] (SD)	F1 score (SD)	Accuracy (SD)	User lift	*P* value
Accelerometer, gyroscope, heart rate					
Condition 1: Watch movie, then walk					
Baseline	0.500 (0.000)	0.348 (0.017)	0.513 (0.015)		
Logistic regression	0.876 (0.085)	0.817 (0.089)	0.818 (0.089)	0.305	<.001
Random forest	0.923 (0.059)	0.854 (0.073)	0.854 (0.073)	0.342	<.001
Condition 2: Listen to music, then walk					
Baseline	0.500 (0.000)	0.342 (0.007)	0.508 (0.006)		
Logistic regression	0.812 (0.081)	0.748 (0.071)	0.748 (0.071)	0.240	<.001
Random forest	0.887 (0.046)	0.806 (0.047)	0.806 (0.047)	0.298	<.001
Condition 3: Listen to music while walking					
Baseline	0.500 (0.000)	0.356 (0.031)	0.520 (0.027)		
Logistic regression	0.900 (0.096)	0.849 (0.107)	0.849 (0.107)	0.329	<.001
Random forest	0.948 (0.057)	0.890 (0.081)	0.891 (0.080)	0.371	<.001
Accelerometer, heart rate					
Condition 1: Watch movie, then walk					
Baseline	0.500 (0.000)	0.348 (0.017)	0.513 (0.015)		
Logistic regression	0.809 (0.105)	0.752 (0.099)	0.753 (0.099)	0.240	<.001
Random forest	0.891 (0.081)	0.821 (0.090)	0.822 (0.089)	0.309	<.001
Condition 2: Listen to music, then walk					
Baseline	0.500 (0.000)	0.342 (0.007)	0.508 (0.006)		
Logistic regression	0.729 (0.070)	0.674 (0.055)	0.675 (0.055)	0.167	<.001
Random forest	0.847 (0.046)	0.768 (0.045)	0.769 (0.045)	0.261	<.001
Condition 3: Listen to music while walking					
Baseline	0.500 (0.000)	0.356 (0.031)	0.520 (0.027)		
Logistic regression	0.876 (0.095)	0.821 (0.106)	0.821 (0.106)	0.301	<.001
Random forest	0.933 (0.067)	0.871 (0.088)	0.871 (0.088)	0.351	<.001
Accelerometer					
Condition 1: Watch movie, then walk					
Baseline	0.500 (0.000)	0.348 (0.017)	0.513 (0.015)		
Logistic regression	0.786 (0.097)	0.726 (0.089)	0.727 (0.089)	0.215	<.001
Random forest	0.847 (0.076)	0.773 (0.077)	0.774 (0.077)	0.261	<.001
Condition 2: Listen to music, then walk					
Baseline	0.500 (0.000)	0.342 (0.007)	0.508 (0.006)		
Logistic regression	0.708 (0.056)	0.657 (0.047)	0.658 (0.047)	0.150	<.001
Random forest	0.783 (0.051)	0.712 (0.042)	0.713 (0.042)	0.205	<.001
Condition 3: Listen to music while walking					
Baseline	0.500 (0.000)	0.356 (0.031)	0.520 (0.027)		
Logistic regression	0.848 (0.086)	0.789 (0.096)	0.790 (0.095)	0.269	<.001
Random forest	0.899 (0.066)	0.825 (0.080)	0.825 (0.079)	0.305	<.001

[a]AUC: area under the curve.

We used the user lift framework to quantify whether a personal model was better than a personal baseline for each user [26]. We calculated the user lift as the difference in the accuracy of the personal classifier and the personal baseline (classifier accuracy–personal baseline accuracy). We used the nonparametric permutation test to determine whether user lifts had a mean greater than 0 (see Table 4). Figure 3 shows the calculated user lift for each participant using random forest model and logistic regression. We included this figure because average user lift can obscure the presence of negative user lift for some participants. Using features extracted from the accelerometer, gyroscope, and heart rate data resulted in the highest accuracies. Using only features from accelerometer data resulted in lower accuracies. Overall for the personal models, the average user lift was greater than 0 for all conditions, indicating that the trained personal models outperformed the baseline.

Happy Versus Neutral Versus Sad

Figure 4 shows the distribution of accuracies of personal models for the three-class classification task of predicting happy-neutral-sad emotional states. We used all features (acceleration, angular velocity, and heart rate) for classification. Although personal models on average outperformed the baseline, accuracies are lower than those achieved when predicting only happy versus sad. Because the number of samples for each class is approximately the same, the baseline predicting the majority class is able to classify correctly only about a third of testing samples. See Table 5 for user lift results. Personal models outperformed personal baselines, but overall accuracy was lower than binary classification of happy versus sad.

Emotion Cross-Validation

We conducted an experiment to assess the effect of neighborhood bias in evaluation of our models using random cross-validation. In this experiment, we conducted 10-fold cross-validation for each personal model, but the testing fold that was held out during each iteration held out a contiguous happy data block or a contiguous sad data block. The goal was to determine with higher confidence whether classifiers were learning patterns associated with emotions, as opposed to just learning to distinguish between different walking periods. In addition, this type of validation takes into consideration neighborhood bias, which can lead to overly optimistic performance estimates [24]. Results (see Figure 5 and Table 6) show that accuracies across all conditions dropped compared with accuracies when using random cross-validation. Personal models outperformed personal baselines but overall accuracy was poor.

Figure 3. The user lift for personal models per condition. The random forest user lift is calculated as (random forest accuracy – baseline accuracy) and the logistic regression user lift is calculated as (logit accuracy – baseline accuracy). The personal models achieve higher accuracies than the personal baseline models.

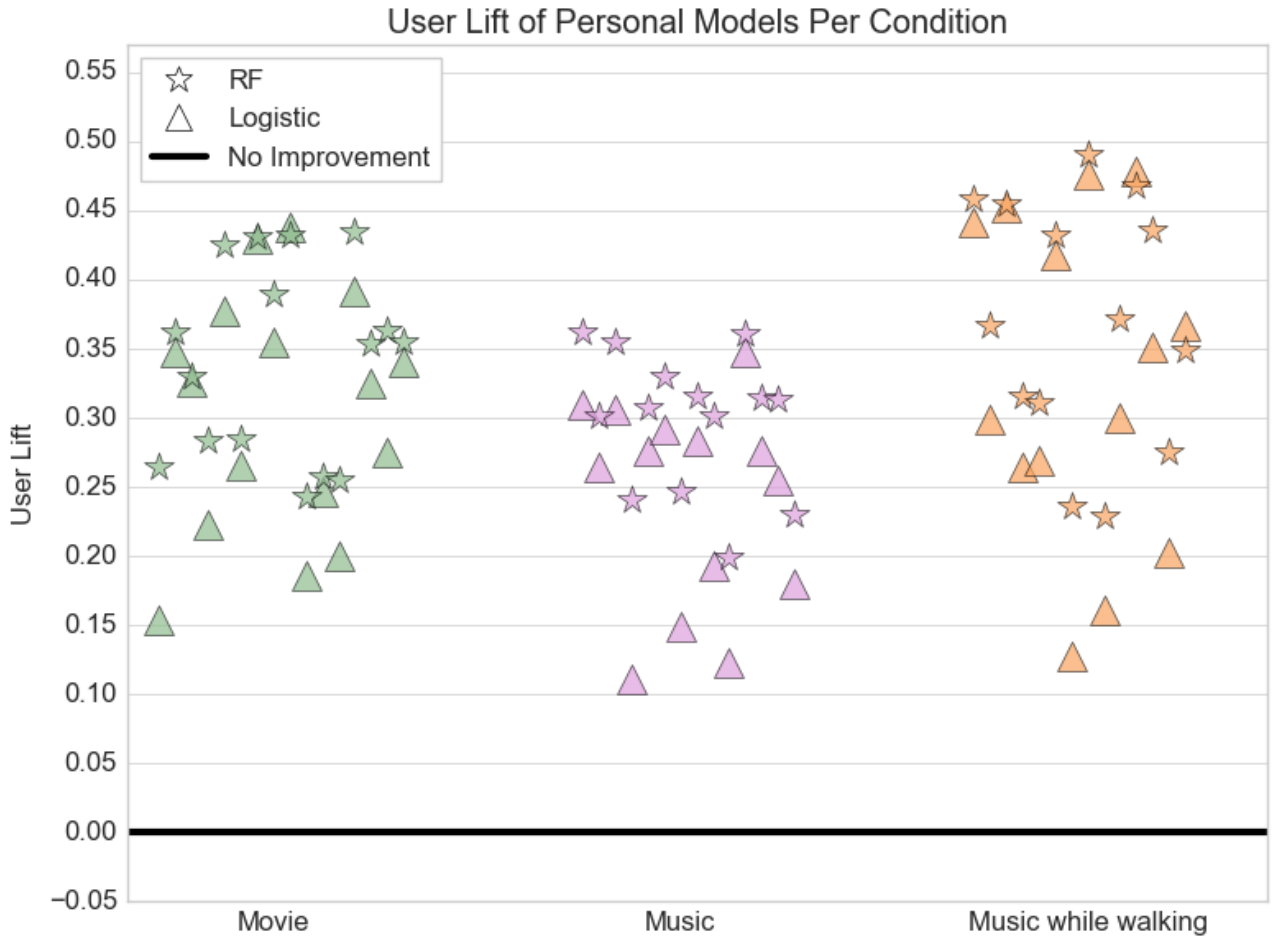


Figure 4. Classification accuracies for participants divided by conditions for the recognition of happiness, sadness, and neutral emotional states. The lower accuracies when recognizing the neutral emotional state indicates that the neutral walking data does have more similarities to the happy and sad walking data, which may indicate the need for additional features. RF: random forest.

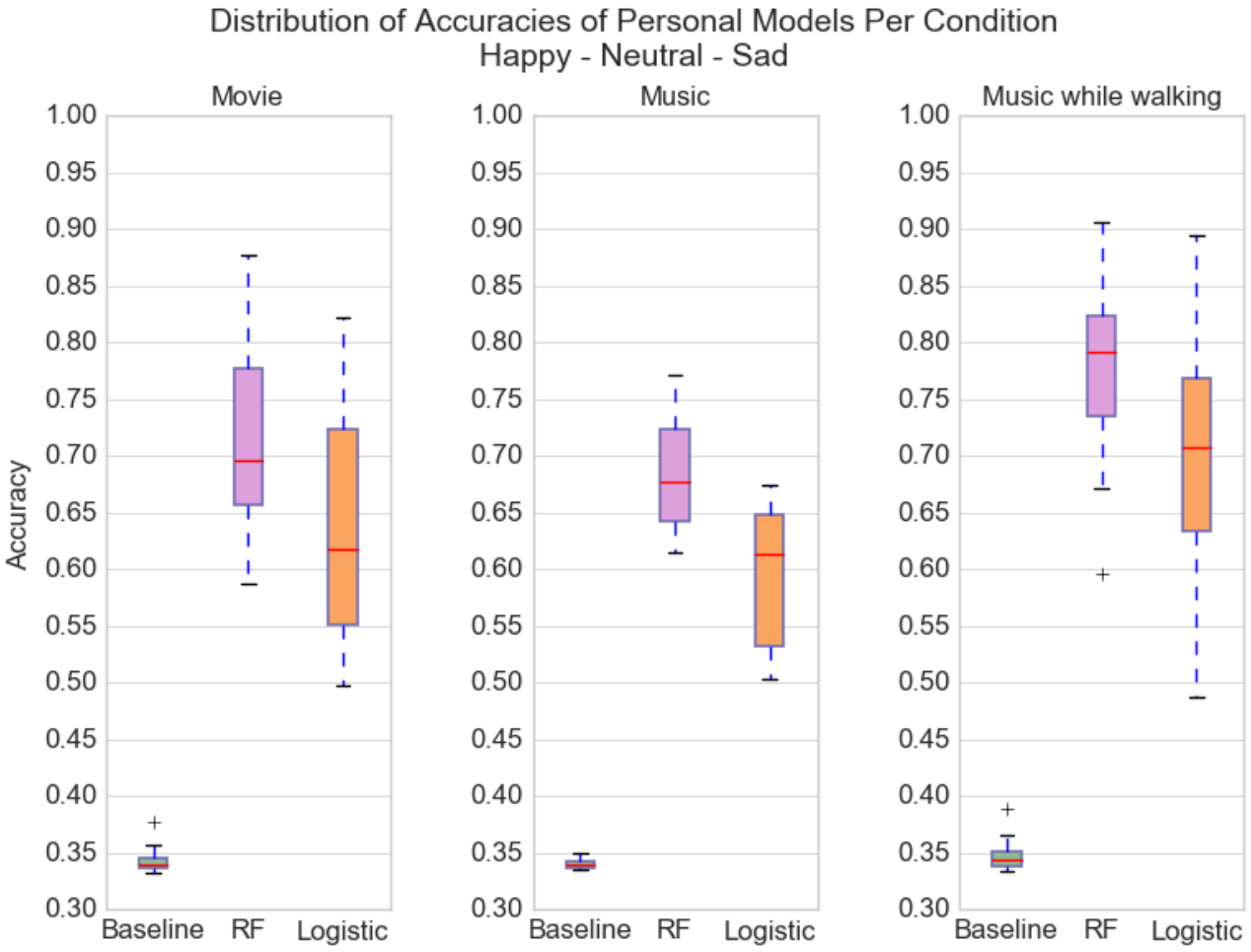


Table 5. Average user lift and average personal model accuracy per condition for the three-class classification task of predicting happy-neutral-sad.

Model	F1 score (SD)	Accuracy (SD)	User lift	*P* value
Condition 1: Watch movie, then walk				
Baseline	0.175 (0.010)	0.343 (0.011)		
Logistic regression	0.632 (0.103)	0.635 (0.103)	0.292	<.001
Random forest	0.722 (0.090)	0.723 (0.090)	0.380	<.001
Condition 2: Listen to music, then walk				
Baseline	0.173 (0.004)	0.340 (0.004)		
Logistic regression	0.591 (0.062)	0.594 (0.061)	0.254	<.001
Random forest	0.684 (0.048)	0.685 (0.047)	0.345	<.001
Condition 3: Listen to music while walking				
Baseline	0.180 (0.014)	0.348 (0.015)		
Logistic regression	0.709 (0.113)	0.711 (0.113)	0.363	<.001
Random forest	0.781 (0.087)	0.782 (0.087)	0.434	<.001

Figure 5. Boxplot of classification accuracies for participants divided by conditions. The results are for 10-fold cross-validation, with each fold in the training data consisting of contiguous windows from both happy and walking data, and the held-out test fold consisting of contiguous windows from either the happy or the sad walking data. RF: random forest.

Table 6. Average user lift and average personal model accuracy per condition.

Model	F1 score (SD)	Accuracy (SD)	User Lift	*P* value
Condition 1: Watch movie, then walk				
Baseline	0.031 (0.121)	0.031 (0.121)		
Logistic regression	0.787 (0.104)	0.682 (0.139)	0.650	<.001
Random forests	0.763 (0.112)	0.651 (0.146)	0.620	<.001
Condition 2: Listen to music, then walk				
Baseline	0.000 (0.000)	0.000 (0.000)		
Logistic regression	0.705 (0.099)	0.575 (0.115)	0.575	<.001
Random forests	0.678 (0.105)	0.543 (0.118)	0.543	<.001
Condition 3: Listen to music while walking				
Baseline	0.036 (0.129)	0.036 (0.129)		
Logistic regression	0.812 (0.140)	0.723 (0.179)	0.688	<.001
Random forests	0.815 (0.148)	0.731 (0.185)	0.695	<.001

Table 7. Accuracy scores for leave-one-user-out cross-validation.

Model	AUC[a] (SD)	F1 score (SD)	Accuracy
Condition 1: Watch movie, then walk			
Baseline	0.500 (0.000)	0.342 (0.021)	0.508 (0.018)
Logistic regression	0.539 (0.137)	0.461 (0.112)	0.515 (0.090)
Condition 2: Listen to music, then walk			
Baseline	0.500 (0.000)	0.332 (0.011)	0.499 (0.010)
Logistic regression	0.539 (0.084)	0.467 (0.061)	0.519 (0.059)
Condition 3: Listen to music while walking			
Baseline	0.500 (0.000)	0.323 (0.034)	0.490 (0.032)
Logistic regression	0.510 (0.173)	0.476 (0.092)	0.505 (0.082)

[a]AUC: area under the curve.

However, the performance of models remains higher than personal baselines with the exception of a few users. Only a quarter of baseline models under Condition 1 and Condition 3 achieved accuracies ranging from 0 to 0.5; the rest have accuracies of 0. This is expected because a baseline model predicted on the majority class will achieve an accuracy of 0 when tested on a contiguous block of the opposite class.

We conclude that for at least half the participants in Condition 1 (movie) and Condition 3 (music while walking), models are likely learning patterns associated with sad and happy emotions. In addition, high accuracies indicate that model performance is not a result of neighborhood bias [24].

Generalizing Across Users

We conducted leave-one-user-out cross-validation to assess how well a model trained on data from certain users would be able to generalize to a user for whom no data are available. We compared both the logistic regression and random forest model. However, random forest models performed similarly or worse than logistic regression; therefore, we only discussed results of the best performing logistic regression compared against the baseline (see Table 7). Logistic regression performed poorly across all conditions, showing that using data from different users to do emotion recognition on a different user is not possible with current features and logistic regression. Low accuracies across all conditions show that the behavior from user to user varies considerably, even when performing a similar action. Owing to the small number of users per condition (<18), data may not be enough to make accurate predictions for users not included in the training set [24]. However, it also highlights a limitation in our modeling approach, in that different features or more advanced models may be necessary to generalize across users. Ideally, deployment of an app should include an initial data collection and calibration phase, which can be used to build a high accuracy personal model for each user.

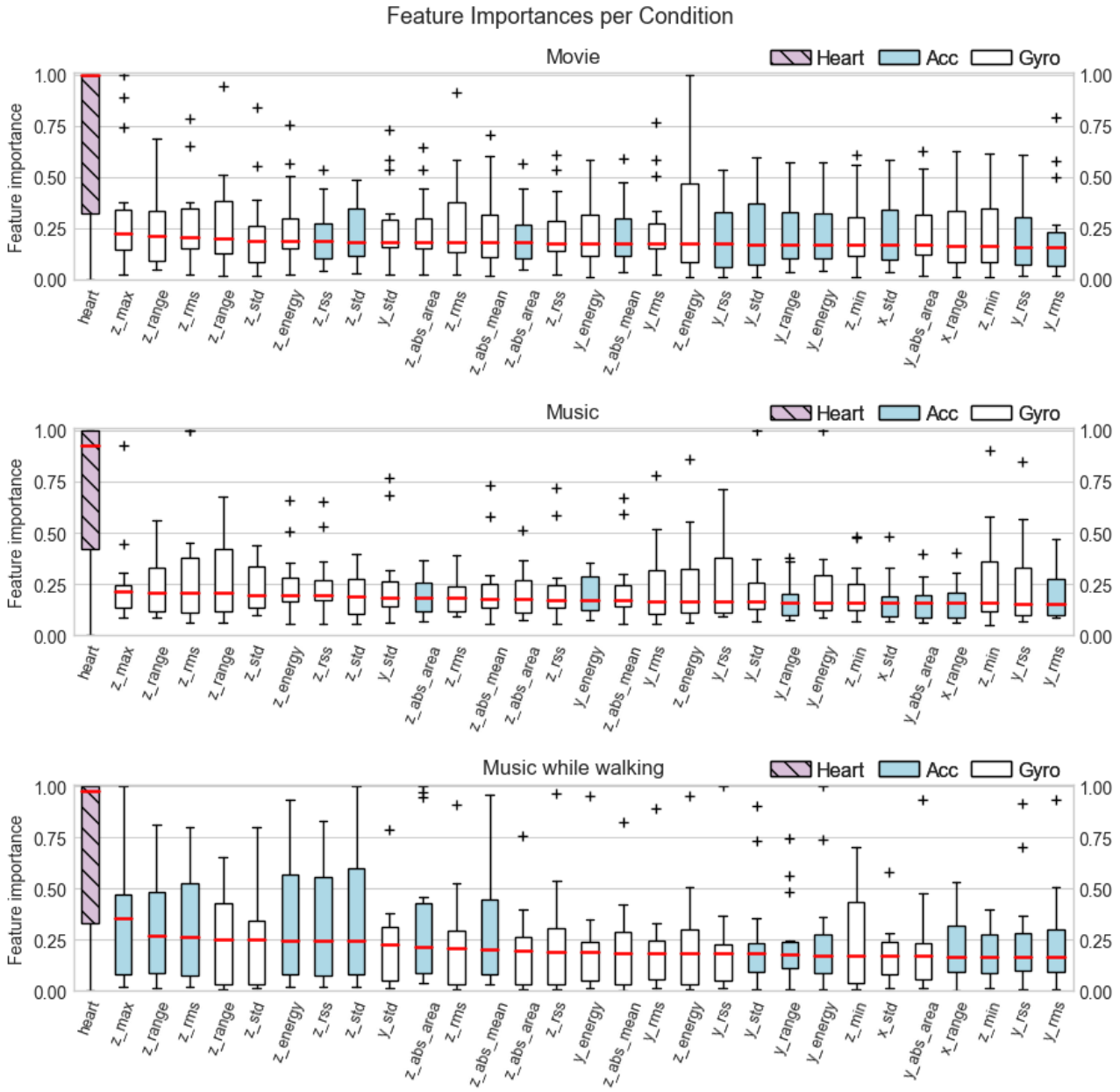
Model Interpretability

We address model interpretability, that is, how models are able to differentiate between emotions, by examining information gain of features. Random forest models can be interpreted by examining feature importances, and logistic regression can be interpreted by the sign and value of the coefficients. Random forest models outperformed logistic regression in our results; therefore, we limit our analysis to feature importances of random forest models.

Because we are building personal models, features that might be important for one user may be less important for another user. To show this, we plotted the distribution of feature importance values for each feature across all users using boxplots, as seen in Figure 6. Boxplots are sorted by median and we included only the top 30 features for visibility with the trend of the remaining features being about the same. To obtain feature importances for each user, we computed the mean feature importance for each feature in cross-validation folds and divided each feature by the maximum feature importance value. Thus, a value of 1.0 indicates that a feature was the most important among all the features.

A compact boxplot indicates that the feature has similar importance across all users. On the other hand, a boxplot with a large spread indicates that the feature is important for some users but less important for other users. For all conditions, heart rate was the most important feature. In fact, for Condition 1 (movie), heart rate was the most important feature for at least half of users (median=1.0). The rest of the features have distributions with smoothly decreasing medians but with heart rate being the only feature with a clear difference from other features.

Figure 6. Distribution of feature importances per feature for all personal models. Acc: accelerometer; Gyro: gyroscope.



Discussion

Behavioral Response to Stimuli

Participants reported feeling less negative affect after watching sad videos or after listening to sad music. This is contrary to the other condition (listening to music while walking) in which participants reported feeling more negative during happy stimuli compared with the neutral ones. Our findings suggest that the walking activity after experiencing a stimulus is useful to alleviate negative mood, similar to that reported [27,28], but not while experiencing stimuli. One reason for this is that participants were focused on the song, and possibly, the change between music types creates resentment or unhappiness. Some studies suggest that people may prefer sad music [29,30], which may influence participants' response toward stimuli. However, a subset of 10 participants reported liking the sad stimulus the least compared with happy and neutral stimuli, suggesting that this is not a case of liking sad music more than others. This personal preference self-report further adds credence to PANAS results that walking is useful in alleviating negative mood.

From heart rate data, our participants did not experience any significant difference in heart rate between emotions. One possible explanation is that walking itself is a vigorous activity compared with standing still; thus, the brief exposure to emotional stimulus may not have been captured holistically. The other possible explanation is that both emotions were equally successful in evoking their emotional states; therefore, there was a nonsignificant difference between them. Nonetheless, data from PANAS suggest that it is likely the latter because participants reported experiencing a difference between positive and negative states.

Classifiers for Emotion Recognition

High accuracies achieved across all users for classification of happy versus sad emotional states provide further evidence for

the hypothesis that movement sensor data can be used for emotion recognition. To build personal models, we used statistical features that are computationally cheap, which would make it feasible to deploy a smart watch or a mobile phone app that can track emotions from movement sensor data without taxing the smart watch or mobile phone processor.

Using only accelerometer data for emotion recognition resulted in mean AUC values of at least 71% for all conditions. The combination of accelerometer data features and heart increased the overall performance of models to a mean AUC of 73%. The use of accelerometer, heart rate, and gyroscope features increased the mean AUC to 81%. This provides a strong motivation to use gyroscope and heart rate data in applications attempting to infer emotional states from movement data, especially given that application programming interfaces of mobile phones and smart watches make it easy to retrieve gyroscope and heart rate. In addition, the high importance of the heart rate feature in random forest models ought to encourage developers to use heart rate data from a smart watch for emotion recognition.

When comparing the classification results using features extracted from all sensor data on classification of happy versus sad emotions, we achieved high-fidelity emotion recognition models with an accuracy of ≥80% for 62.5% (55/88) of the personal models, average-fidelity models with an accuracy between 70% and 80% for 27.3% (24/88) of the personal models, and low-fidelity models with an accuracy of <70% for 10.2% (9/88) of the personal models. These results are encouraging. However, they also indicate that further work is needed to achieve consistent results across different users and accuracies closer to 100%. For example, this could be achieved by extracting additional features, using a more complex classifier, or by collecting more data for training and testing personal models. Lastly, our results on emotion cross-validation highlight that personal models for about half the participants are learning features that capture emotions.

Limitations

Previous studies have utilized a contrast experimental paradigm to manipulate the following participants' moods: positive versus negative mood [2]; negative or neutral [31]; positive, negative, and neutral [32,33] using music or avatars. Past research findings indicate that negative moods tend to reduce mood recovery and a slower response for accurately identifying other emotional expressions [20,31]. Although these user studies did not apply to emotion recognition from sensor data obtained from a smart watch, we did not address issues, such as reduced mood recovery, for participants who were shown the sad stimulus first; however, we did perform counterbalancing for our stimuli on our participants.

The integrity of sensor data is a concern. For Conditions 1 and 2, participants were primed with audio and audiovisual stimuli for a few minutes, but beyond PANAS scores, we do not have other means to indicate that the stimulus had the intended effect. Furthermore, the effect of the stimulus on participants is questionable given that participants were not emotionally invested in movie and music clips that were shown. Personal models do distinguish at high accuracies between features extracted from happy, sad, and neutral emotions, but we do not know for certain that happy data is truly associated with a "happy" emotional state in users. In general, given that the mixed-design study consisted of 3 conditions, 50 participants is a small sample size.

From a modeling and data analysis point of view, the amount of data collected was small. Hence, this limits the training and validation of classifiers. Although personal models yielded high accuracies for many users, for other users, the results were slightly better than random guessing. Finally, we did not consider more flexible modeling approaches, such as using a time-aware model or using a neural network trained on raw sensor data, instead of extracting features from sliding windows.

The personal models we built are naïve, in that each window is an independent sample. Therefore, a model could potentially predict happy-sad-happy for 3 consecutive one-second windows, which is unrealistic as a user is not likely to go from happy to sad and back to happy in a matter of 3 seconds. This limitation of our modeling approach will be addressed in future work.

Comparison With Prior Work

Our work is closest to the work reported previously [10,11]. In [10], the details of the design study are omitted, including the choice of videos and procedures. A limitation in a previous study [10] is that data was collected from two mobile phones, one strapped to the wrist and one strapped to the ankle of participants. In a previous study [11], 123 participants were recruited (twice the size of our sample), and smart bracelets were used for data collection with participants wearing a smart bracelet on the wrist and another smart bracelet on the ankle (with the latter violating ecological validity). We achieved accuracies comparable to those reported in another study [11], using only data from one smart watch on participants' wrists and without relying on data from other body locations. Our work also differs in that we focus on happy and sad emotional states, whereas in a study [11], researchers focused on happy and angry emotional states. In contrast to prior work, we performed more rigorous testing by including emotion cross-validation and by extracting features from an accelerometer, a gyroscope, and heart rate sensors.

In contrast to emotion prediction based on typing behavior [5,6], mobile phone usage [7,8], and mobile phone speech recordings [9], we focused on movement data and heart rate data. The EmotionSense system does use accelerometer data to determine whether a user is moving but not for emotion recognition [9].

Conclusions and Future Work

Our findings suggest that emotional expression is transparent even in automatic functions such as walking gait. This finding is interesting in that healthy young adults typically do not report large differences in their emotional state, unlike some clinical groups [34]. These findings also validate our methodological approach with respect to priming the emotional state and the subsequent modeling using machine-learning algorithms.

Many studies have focused on face and voice modalities, but recent studies have shown that we tend to adopt different body postures and gaits as a reflection of our emotions and that these

postures and gaits are just as easily recognized by others, indicating that walking gait is a form of social signal. However, the emotional behavioral response is only evident after experiencing the stimulus on its own or while experiencing both together (eg, listening to music while walking). Nonetheless, our findings provide further knowledge in the field of social communication, particularly in specific clinical conditions. The unobtrusive wearable is a good complement for collecting data and for providing biofeedback and interventions for emotional regulation. Recent studies have started analyzing the possibility of using wearables to provide more readily available treatment for patients and provide feedback to clinicians to cater to their needs [34-36]. Benefits of using these wearables, particularly in identifying emotional states, are useful for diagnosis or monitoring specific clinical conditions, such as social anxiety and borderline personality disorder. Although most research is focused on getting patients to self-rate their moods, having actigraph data and walking patterns will complement the information necessary for clinicians. Other than for a clinical population, this type of information is also useful for vulnerable populations (eg, older adults) experiencing some emotional distress and social isolation [37]. Future studies should look into the duration of having on such wearables (over 24-hour cycles) and duration in experiencing stimuli (acute or chronic experiences).

Acknowledgments

The authors wish to thank all the volunteers who participated in our user study and Elisa Roberti and Magdalene Rose for assisting in data collection. This research is funded by the Lancaster Small Grant Scheme awarded to JQ and MY (SGSSL-FST-DCIS-0015-02). JQ was affiliated with the Department of Computer Science and Engineering at the University of Nevada, Reno, at the time that this work was completed and is currently affiliated with the Australian Institute of Health Innovation at Macquarie University.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Means and SDs in brackets for positive and negative affect scores for each emotion.

[PDF File (Adobe PDF File), 22KB - mental_v5i3e10153_app1.pdf]

References

1. Ekman P. Emotional and Conversational Nonverbal Signals. In: Larrazabal JM, Perez Miranda LA, editors. Language, Knowledge, and Representation: Proceedings of the Sixth International Colloquium on Cognitive Science (ICCS-99) (Philosophical Studies Series). Berlin: Springer; 2012:39-50.
2. Michalak J, Troje NF, Fischer J, Vollmar P, Heidenreich T, Schulte D. Embodiment of sadness and depression--gait patterns associated with dysphoric mood. Psychosomatic Medicine 2009 Jun;71(5):580-587. [doi: 10.1097/PSY.0b013e3181a2515c] [Medline: 19414617]
3. Zeng Z, Pantic M, Roisman GI, Huang TS. A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans Pattern Anal Mach Intell 2009 Jan;31(1):39-58. [doi: 10.1109/TPAMI.2008.52] [Medline: 19029545]
4. El Ayadi M, Kamel MS, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition 2011 Mar;44(3):572-587. [doi: 10.1016/j.patcog.2010.09.020]
5. Lee H, Choi YS, Lee S, Park IP. Towards unobtrusive emotion recognition for affective social communication. 2012 Presented at: IEEE Consumer Communications and Networking Conference (CCNC); 14-17 Jan. 2012; Las Vegas, NV, USA. [doi: 10.1109/CCNC.2012.6181098]
6. Shapsough S, Hesham A, Elkhorazaty Y, Zualkernan I, Aloul F. Emotion recognition using mobile phones. 2016 Presented at: IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom). .?6; 14-17 September 2016; Munich, Germany p. 2016. [doi: 10.1109/HealthCom.2016.7749470]
7. Chittaranjan G, Blom J, Gatica-Perez D. Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones. 2011 Presented at: 2011 15th Annual International Symposium on Wearable Computers (ISWC); June 12, 2011 to June 15, 2011; San Francisco, CA. [doi: 10.1109/ISWC.2011.29]
8. Drake G, Csipke E, Wykes T. Assessing your mood online: acceptability and use of Moodscope. Psychol Med 2013 Jul;43(7):1455-1464 [FREE Full text] [doi: 10.1017/S0033291712002280] [Medline: 23149120]
9. Rachuri K, Musolesi M, Mascolo C, Rentfrow P, Longworth C, Aucinas A. EmotionSense: A Mobile Phones Based Adaptive Platform for Experimental Social Psychology Research. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10). USA: ACM; 2010 Presented at: Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10); 2010; Copenhagen, Denmark p. 281-290. [doi: 10.1145/1864349.1864393]

10. Cui L, Li S, Zhu T. Emotion Detection from Natural Walking. In: Human Centered Computing Internet.: Springer, Cham; 2016 Presented at: International Conference on Human Centered Computing; 7-9 January 2016; Colombo, Sri Lanka URL: https://link.springer.com/chapter/10.1007/978-3-319-31854-7_3 [doi: 10.1007/978-3-319-31854-7_3]
11. Zhang Z, Song Y, Cui L, Liu X, Zhu T. Emotion recognition based on customized smart bracelet with built-in accelerometer. PeerJ 2016;4:e2258 [FREE Full text] [doi: 10.7717/peerj.2258] [Medline: 27547564]
12. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using Smartphones to Monitor Bipolar Disorder Symptoms: A Pilot Study. JMIR Ment Health 2016 Jan 06;3(1):e2 [FREE Full text] [doi: 10.2196/mental.4560] [Medline: 26740354]
13. Sarker H, Hovsepian K, Chatterjee S, Nahum-Shani I, Murphy S, Spring B, et al. From Markers to Interventions: The Case of Just-in-Time Stress Intervention. In: Mobile Health. New York City: Springer, Cham; 2017:411-433.
14. Zheng YL, Ding XR, Poon CCY, Lo BPL, Zhang H, Zhou XL, et al. Unobtrusive sensing and wearable devices for health informatics. IEEE Trans Biomed Eng 2014 May;61(5):1538-1554. [doi: 10.1109/TBME.2014.2309951] [Medline: 24759283]
15. Appelhans B, Luecken L. Heart rate variability as an index of regulated emotional responding. Review of General Psychology 2006;10(3):229-240. [doi: 10.1037/1089-2680.10.3.229]
16. Quiroz J, Yong M, Geangu E. Emotion-recognition Using Smart Watch Accelerometer Data: Preliminary Findings. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous ComputingProceedings of the 2017 ACM International Symposium on Wearable Computers Internet. USA: ACM; 2017 Presented at: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers; 11/09/17 to 15/09/17; Maui, Hawaii, USA. [doi: 10.1145/3123024.3125614]
17. Gross J, Levenson R. Emotion elicitation using films. Cognition & Emotion 1995 Jan;9(1):87-108. [doi: 10.1080/02699939508408966]
18. Bartolini E. Eliciting emotion with film: development of a stimulus set degree. In: WesScholar. Connecticut: Wesleyan University; 2011:84.
19. Schaefer A, Nils F, Sanchez X, Philippot P. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition & Emotion 2010 Nov;24(7):1153-1172. [doi: 10.1080/02699930903274322]
20. Mitterschiffthaler M, Fu C, Dalton J, Andrew C, Williams S. A functional MRI study of happy and sad affective states induced by classical music. Hum Brain Mapp 2007 Nov;28(11):1150-1162. [doi: 10.1002/hbm.20337] [Medline: 17290372]
21. Watson D, Clark L, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology 1988 Jun;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063] [Medline: 3397865]
22. Anguita D, Ghio A, Oneto L, Ortiz R, Luis J. A Public Domain Dataset for Human Activity Recognition Using Smartphones. In: Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2013 Presented at: 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning; 24 - 26 April 2013; Bruges p. 437-442.
23. Reyes-Ortiz J, Oneto L, Samà A, Parra X, Anguita D. Transition-Aware Human Activity Recognition Using Smartphones. Neurocomputing 2016 Jan;171:754-767. [doi: 10.1016/j.neucom.2015.07.085]
24. Hammerla N, Plötz T. Let's (Not) Stick Together: Pairwise Similarity Biases Cross-validation in Activity Recognition. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. USA: ACM; 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 07 - 11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2807551]
25. Lockhart J, Weiss G. Limitations with Activity Recognition Methodology & Data Sets. In: Proceedings of the 2014 ACM International Joint Conference on PervasiveUbiquitous Computing: Adjunct Publication Internet. New York, NY. USA: ACM; 2014 Presented at: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication; September 13 - 17, 2014; Seattle, Washington. [doi: 10.1145/2638728.2641306]
26. DeMasi O, Kording K, Recht B. Meaningless comparisons lead to false optimism in medical machine learning. PLoS One 2017;12(9):e0184604 [FREE Full text] [doi: 10.1371/journal.pone.0184604] [Medline: 28949964]
27. Bernstein E, McNally RJ. Acute aerobic exercise helps overcome emotion regulation deficits. Cogn Emot 2017 Jun;31(4):834-843. [doi: 10.1080/02699931.2016.1168284] [Medline: 27043051]
28. Penedo F, Dahn J. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry 2005 Mar;18(2):189-193. [Medline: 16639173]
29. Huron D. Why is Sad Music Pleasurable? A Possible Role for Prolactin. Musicae Scientiae 2011 Jul 13;15(2):146-158. [doi: 10.1177/1029864911101500202]
30. Sachs M, Damasio A, Habibi A. The pleasures of sad music: a systematic review. Front Hum Neurosci 2015;9:404 [FREE Full text] [doi: 10.3389/fnhum.2015.00404] [Medline: 26257625]
31. Veenstra L, Schneider I, Koole S. Embodied mood regulation: the impact of body posture on mood recovery, negative thoughts, and mood-congruent recall. Cogn Emot 2017 Nov;31(7):1361-1376. [doi: 10.1080/02699931.2016.1225003] [Medline: 27626675]

32. Janssen D, Schöllhorn W, Lubienetzki J, Fölling K, Kokenge H, Davids K. Recognition of Emotions in Gait Patterns by Means of Artificial Neural Nets. J Nonverbal Behav 2008 Jan 11;32(2):79-92. [doi: 10.1007/s10919-007-0045-3]
33. Schneider S, Christensen A, Häußinger FB, Fallgatter A, Giese M, Ehlis AC. Show me how you walk and I tell you how you feel - a functional near-infrared spectroscopy study on emotion perception based on human gait. Neuroimage 2014 Jan 15;85 Pt 1:380-390. [doi: 10.1016/j.neuroimage.2013.07.078] [Medline: 23921096]
34. Arribas I, Saunders K, Goodwin G, Lyons T. ArXiv170707124 Stat Internet. 2017. A signature-based machine learning model for bipolar disorder and borderline personality disorder URL: http://arxiv.org/abs/1707.07124 [accessed 2018-07-20] [WebCite Cache ID 713EOWFTn]
35. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
36. Kormilitzin A, Saunders K, Harrison P, Geddes J, Lyons T. ArXiv170801206 Stat. 2017. Detecting early signs of depressive and manic episodes in patients with bipolar disorder using the signature-based model URL: http://arxiv.org/abs/1708.01206 [accessed 2018-07-20] [WebCite Cache ID 713ESpPEb]
37. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci U S A 2013 Apr 09;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]

Abbreviations

AUC: area under the curve

Edited by R Calvo, M Czerwinski, J Torous, G Wadley; submitted 15.02.18; peer-reviewed by U Lee, D Snider; comments to author 06.03.18; revised version received 01.05.18; accepted 19.05.18; published 08.08.18.

Please cite as:
Quiroz JC, Geangu E, Yong MH
Emotion Recognition Using Smart Watch Sensor Data: Mixed-Design Study
JMIR Ment Health 2018;5(3):e10153
URL: http://mental.jmir.org/2018/3/e10153/
doi:10.2196/10153
PMID:30089610

Original Paper

A Process Evaluation of a Web-Based Mental Health Portal (WalkAlong) Using Google Analytics

Michael Jae Song[1], MPH; John Ward[1,2], PhD; Fiona Choi[1,2], PhD; Mohammadali Nikoo[1,2], MD; Anastasia Frank[1,2], MPH; Farhud Shams[1], BSc (Eng); Katarina Tabi[1,3], MSc; Daniel Vigo[4], MD, DrPH; Michael Krausz[1,2,5], MD, PhD

[1]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

[2]Centre for Health Evaluation and Outcome Sciences, St Paul's Hospital, Vancouver, BC, Canada

[3]Department of Pharmacology, Faculty of Medicine, Masaryk University, Brno, Czech Republic

[4]Centre for Applied Research in Mental Health and Addictions, Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[5]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Michael Jae Song, MPH
Department of Psychiatry
University of British Columbia
430-5950 University Boulevard
David Strangway Building
Vancouver, BC, V6T1Z3
Canada
Phone: 1 778 899 5205
Email: michaeljaesong@gmail.com

Abstract

Background: Despite the increasing amount of research on Web-based mental health interventions with proven efficacy, high attrition rates decrease their effectiveness. Continued process evaluations should be performed to maximize the target population's engagement. Google Analytics has been used to evaluate various health-related Web-based programs and may also be useful for Web-based mental health programs.

Objective: The objective of our study was to evaluate WalkAlong.ca, a youth-oriented mental health web-portal, using Google Analytics to inform the improvement strategy for the platform and to demonstrate the use of Google Analytics as a tool for process evaluation of Web-based mental health interventions.

Methods: Google Analytics was used to monitor user activity during WalkAlong's first year of operation (Nov 13, 2013-Nov 13, 2014). Selected Google Analytic variables were overall website engagement including pages visited per session, utilization rate of specific features, and user access mode and location.

Results: The results included data from 3076 users viewing 29,299 pages. Users spent less average time on Mindsteps (0 minute 35 seconds) and self-exercises (1 minute 08 seconds), which are important self-help tools, compared with that on the Screener tool (3 minutes 4 seconds). Of all visitors, 82.3% (4378/5318) were desktop users, followed by 12.7 % (677/5318) mobile phone and 5.0% (263/5318) tablet users. Both direct traffic (access via URL) and referrals by email had more than 7 pages viewed per session and longer than average time of 6 minutes per session. The majority of users (67%) accessed the platform from Canada.

Conclusions: Engagement and feature utilization rates are higher among people who receive personal invitations to visit the site. Low utilization rates with specific features offer a starting place for further exploration of users in order to identify the root cause. The data provided by Google Analytics, although informative, can be supplemented by other evaluation methods (ie, qualitative methods) in order to better determine the modifications required to improve user engagement. Google Analytics can play a vital role in highlighting the preferences of those using Web-based mental health tools.

(JMIR Ment Health 2018;5(3):e50) doi:10.2196/mental.8594

KEYWORDS

evaluation; Google Analytics; mental health; website

Introduction

As technologies such as internet, mobile phones, and computers have become ubiquitous, Web-based interventions have become one of the major treatment and preventative tools for mental disorders. More than 100 randomized controlled trials have been published to demonstrate the efficacy of internet interventions for psychiatric disorders [1]. These tools have been shown to be effective for a range of mental illnesses including panic disorder, depression, posttraumatic stress disorder (PTSD), perceived stress in schizophrenia, stress, insomnia, and eating disorders [2]. The potential of Web-based mental health interventions to effectively and efficiently treat and prevent mental illnesses has attracted many health care providers and researchers to explore using them as one of the major components of the mental health care system.

Despite their promising benefits, online mental health interventions face problems in engagement [3-5]. In relation to taking a drug, Web-based interventions are more vulnerable to disengagement as they lack close supervision, are easy to discontinue, and have no immediate health benefits [5]. Dropout rates for Web-based interventions are up to 50% in guided interventions and up to 74% in unguided interventions [6,7]. With such high dropout rates, these interventions may provide only limited outcomes regardless of their proven efficacy [8,9].

Engagement is especially challenging in the context of mental health. Dropout from traditional treatment among those with mental illness is already a cause for concern [10]. The chronic nature of mental illness requires frequent assessments in conjunction with long-term management that is tailored to individuals' needs [11]. In addition, disorder-specific features such as symptom severity, emotional distress, and medication side effects have been shown to predict adherence [12,13]. In a technological context, being engaged can be described as "a category of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety or novelty, interactivity, and perceived user control" [14]. In other words, scientists and developers have the responsibility of ensuring not only the efficacy of their intervention but also user engagement. Strategies to increase user engagement may include continual platform improvement and improving outreach or marketing strategies.

A key step to improve engagement is identifying ways to improve intervention uptake in real-world settings through process evaluation [5]. A process evaluation is a type of assessment that determines whether program activities have been implemented as intended [15]. The goal is to inform strategies toward achieving optimal engagement and effectiveness of an intervention [16-18]. By understanding how users engage with the intervention, such as the specific pages visited or how long they used the website, process evaluation can inform the adaptation of the intervention in order to maximize user exposure to the tools and the knowledge available in Web-based interventions [18]. Ideally, this would involve a mixed-methods approach where both quantitative and qualitative indicators can collectively measure user perceptions or behavior [17]. However, this is not always possible for asynchronous, open-access, Web-based interventions and with limited resources.

In this context, one tool that can be used is Google Analytics, which is an open tool that provides free quantitative data on website usage that can be leveraged for continual website improvements (Figure 1). Although this tool is designed to provide insights from a marketing perspective, numerous variables about the webpage traffic are collected that can inform the process evaluation of Web-based interventions. Indeed, Google Analytics has already been used in health research as part of process evaluation [19-21]. For example, this tool has been used to assess the usage of a website about sexual health [19], an internet-delivered genetics education resource developed for nurses [21], and a Web-based tool to encourage the proper use of antibiotics [16], as well as websites related to osteoporosis and fractures [22], smoking cessation [23], and knowledge translation [24]. These studies have presented various indicators available from Google Analytics to show overall user engagement with their platforms. However, to our knowledge, the use of Google Analytics has not yet been demonstrated for Web-based mental health platforms that provide direct support for youth with mental health challenges. Since Web-based mental health platforms may face challenges with adherence, Google Analytics could be used to better understand user behavior as part of process evaluation and to come up with strategies that would improve adherence.

Figure 1. Google Analytics can be used as a tool for process evaluation by receiving information on user traffic and subsequently informing website improvement. This process can also continue as a cycle for continual improvement of the website.

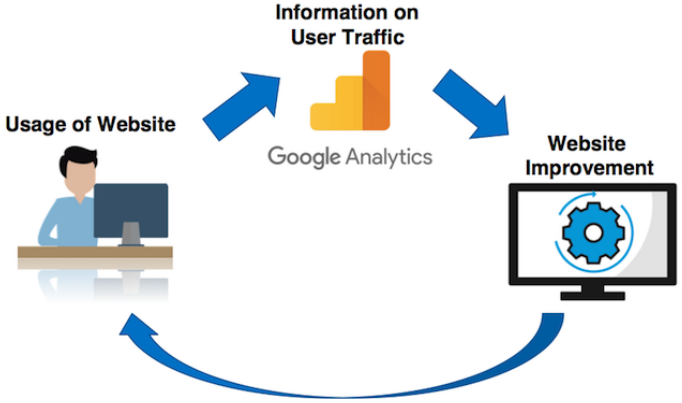
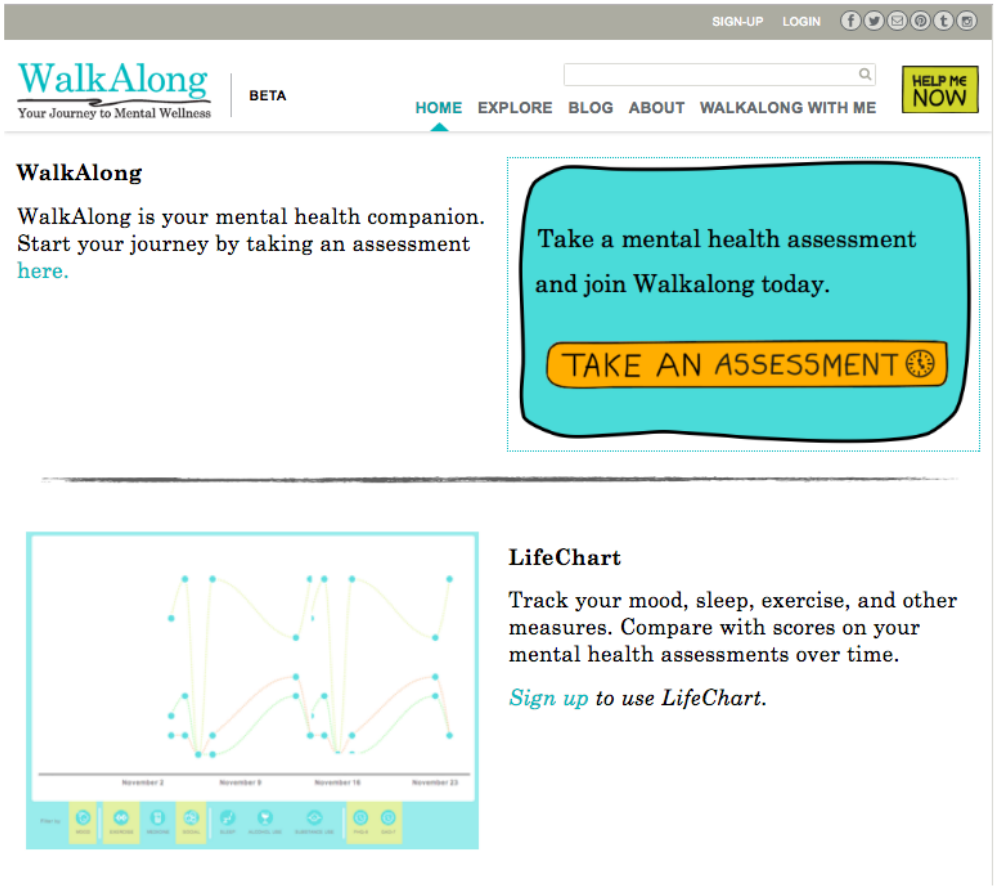


Figure 2. WalkAlong home page.



The goal of this study was to evaluate WalkAlong, a Web-based mental health platform, using Google Analytics. This platform is a youth-oriented mental health Web-portal designed to provide young people with tools and resources required to manage their own mental health (Figure 2; Multimedia Appendix 1). The Web-portal focuses on supporting mood and anxiety disorders and has received funding under Bell Canada's Let's Talk mental health initiative. The portal is freely accessible to all users with an internet connection and contains information, links to resources, and self-help tools including a description and link to MoodGYM, a Web-based resource for depression and anxiety. Self-help tools available directly on the platform include the "Mind Steps" page, which consists of regularly posted short tips for helping users get through the day, and "The Self-Help Exercises" section, displaying a menu with the individual self-help exercise pages. Assessment screeners for depression and anxiety (Patient Health Questionnaire-9 and Generalized Anxiety Disorder 7-item scale) are also available. Users may also create a password-protected, secure account that grants them access to additional resources including the PTSD CheckList—Civilian Version, community resources, and a Life-Chart tracking tool that tracks mood and behaviors.

Our objective was to use Google Analytics as a tool for conducting a process evaluation of the WalkAlong platform. As part of the process evaluation, the following evaluation questions using Google Analytics were asked: (1) How engaged (ie, time spent on the website) are users with the WalkAlong platform? (2) How can the WalkAlong platform be improved to better engage the users? (3) How can the marketing strategy be shaped to engage and reach out to more users? Another objective of this project is to extend the work to a mental health platform from other health interventions and inform website design and marketing strategies to effectively impact user behavior [19,21].

Methods

Google Analytics

Google Analytics was used to access user data over the first year of WalkAlong (Nov 13, 2013-Nov 13, 2014). Focusing on the first year of operation was considered the most appropriate approach in order to capture a snapshot of web traffic following the initial launch. The Google Analytics data do not contain any personally identifiable information and are presented in the form of aggregate data, making it an accessible tool used in research settings without ethical concerns [20,21].

The research team installed Google Analytics by adding a tracking tag for WalkAlong [20]. These tracking tags are snippets of JavaScript code, a computer programming language used to build websites. This code allows collection of various forms of data related to user behavior as soon as the user visits the website. The data can emanate from various avenues such as the URL of the page the user is viewing, the language or the name of the browser, and the device used to access the site. The code also collects information on the nature of the visit such as the contents viewed, length of the session, and channels used to access the platform (eg, Google, direct URL search, social media, and email link). Such information is summarized in a

real-time, interactive dashboard format, which can be accessed by logging in.

Overall Engagement

Several indicators from Google Analytics that would allow inference of a level of engagement were calculated. Such indicators include the number of returning users (n), bounce rate (%), number of pages accessed per session (n), mean session duration (minutes, seconds), and goal conversion rate (%).

The number of returning users refers to the number of sessions visited through the same client id. A high number of returning users has been used as an indicator for a strong level of engagement with the platform [21,25].

The bounce rate is the percentage of only a single page visit during a session. A high bounce rate could indicate minimal exposure to the intervention due to minimal interaction, but it could also indicate users exiting as they have found what they were looking for right away. However, generally, a low bounce rate can be considered indicative of a high overall engagement, especially for a multicomponent platform like WalkAlong [19]. For instance, there are not much available resources that would provide mental health support in the home page of the platform alone as it simply offers an overview. Users will often need to interact with various tools and webpages in order to obtain the necessary information.

The number of pages per session refers to the number of webpages within the platform that the user viewed in a single session, and the mean session duration (minutes, seconds) refers to the mean duration of time the users spent on the platform. There are limitations to ascertaining engagement through these indicators since they allow for multiple interpretations: a high number of pages per session could result from an increased engagement, but it could also result from a superficial exploration of several pages; similarly, a long session duration can result from increased engagement, but it could also result from a user keeping the webpage open while engaging in other irrelevant activities. Nevertheless, despite these caveats, traffic information provides an approximation of the level of exposure the users had with the platform [25].

The goal conversion rate measures the proportion of sessions that achieved a goal out of the total sessions. The goal was predefined as creating an account but can be defined as any activity the web developer or owner chooses (eg, buying a product). As discussed above, users who create an account are able to access more resources than those who do not (anonymous users). Thus, creating an account was assumed to indicate a stronger level of engagement. Overall, a high number of returning users, low bounce rate, high number of pages viewed per session, high mean session duration, and high goal conversion rates collectively translate to an estimate of a strong level of engagement [20,25].

Platform Improvement

Several indicators from Google Analytics that can inform the improvement of the platform were also selected. Indicators of user behavior such as page views, mean duration of visit, and bounce rate when accessing self-help tools (eg, Mindsteps page, Self-Help Exercises page, and Screener) were analyzed. In addition, the most visited pages were observed in terms of their overall entrance rate, exit rate, and bounce rate to understand which tools or pages were most used or viewed. The entrance rate represents a proportion of sessions starting from a given page, while the exit rate represents a proportion of sessions ending from a given page. The information regarding the entrance rate may provide an understanding about which webpage is serving as the first impression for the users, and the exit rate may indicate the point when users felt disengaged or, on the contrary, had adequate information needed for the session.

Google Analytics also provided data on the type of devices used for access. Such information can allow us to consider whether developing a mobile app for WalkAlong would be helpful or not. The three main devices of interest to the current investigation were desktops, tablets, and mobile phones (counted here as mobile devices).

Marketing Strategy

Google Analytics was also used to inform our marketing strategy, with the goal of reaching as many users as possible. At the outset, the research team had reached out to different youth and university organizations, especially around Vancouver. Twitter and a Facebook accounts were also created to spread awareness about the platform. To improve the marketing strategy, the channels used to access the platform were observed. The channels are direct link (ie, typing the web URL directly into a browser); organic search (ie, entry through a search engine); and referrals via another website, via social media, and via email. Understanding which channels are underutilized and which channel results in the highest level of engagement can help improve the marketing strategy. Locations of users from different countries around the world were also observed.

Results

Overall Engagement

The first year of operation for the WalkAlong platform saw a total of 3076 users, amounting to 5318 sessions and 29,299 page views (Figure 3). On average, users visited 5.51 pages per session with an average session duration of 5 minutes 6 seconds. The average bounce rate was 42.9% where users only viewed a single page; 31.7% (976/3076) users created an account (goal completion).

In terms of the frequency of visits, 80% (4259/5318) of sessions came from users visiting less than nine times, indicating a level of disengagement after a certain number of visits (Table 1).

The number of sessions during the study period decreased with increasing number of visits. However, there was a slight increase at the upper end of sessions from high-frequency visits: 5.8% (311/5318) accounted for 26-50 visits and 4.7% (250/5318) accounted for 51-100 visits over the time period. The number of sessions also decreased with longer session durations (Table 2). These results indicate that the majority of sessions or 65.4% (3477/531) resulted in disengagement within the first minute.

Figure 3. WalkAlong overview presented in Google Analytics.

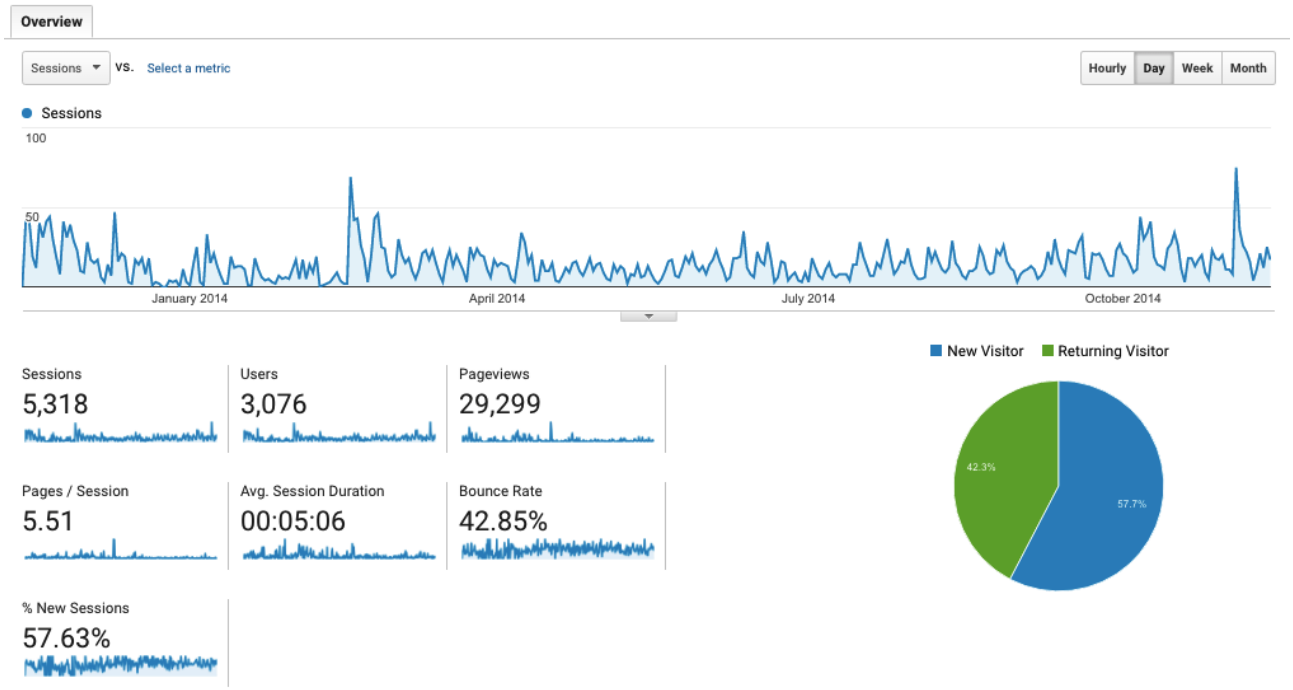


Table 1. Proportion of total sessions and number of visits.

Visits	Sessions (N=5318), n (%)
1	3066 (57.65)
2	550 (10.3)
3	241 (4.5)
4	139 (2.6)
5	100 (1.9)
6	67 (1.3)
7	50 (0.9)
8	46 (0.9)
9-14	180 (3.4)
15-25	205 (3.9)
26-50	311 (5.8)
51-100	250 (4.7)
101-200	112 (2.1)
201+	1 (0.0)

Table 2. Duration of session.

Session duration (in minutes)	Sessions (N=5318), n (%)
≤1	3477 (65.4%)
1-3	527 (9.9%)
3-10	581 (10.9%)
>10	733 (13.8%)

Platform Improvement

Visits to the Mindsteps tool comprised 11.9% (3493/29,299) of total page views, with mean duration spent of 35 seconds. Visits to the Self-Help Exercises page comprised 6.13% (1797/29,299), with mean duration spent of 1 minute 8 seconds. Visits to the Screener comprised only 3.36% (983/29,299) of the total page views, but the mean duration spent on the Screener was 3 minutes 4 seconds. Table 3 presents the entrance and exit rates for the most viewed pages, which included the Self-Help Exercises, Mindsteps, and the Screener. The WalkAlong home page, which acts as the landing page, accounted for 65.6% (3487/5308) of all entries.

A list of devices used by WalkAlong's users to access the site is presented in Table 4, indicating that the platform was accessed mostly via desktops (4378/5318, 82.3%). Furthermore, sessions completed via desktops had a lower bounce rate (39.6%), higher pages per session (6.17), and a higher conversion rate (22.7%) than those completed via other devices.

Marketing Strategy

Direct traffic accounted for the highest proportion (2420/5318, 45.5%) of all visits to the site (Table 5). The combination of high bounce rate (50.4%) and low conversion rate (11.3%) among organic searches suggests that these particular users did not engage much with the site content. Visits via referrals or social media sites had relatively less traffic at 16.0% (849/5318) and 13.5% (717/5318), respectively, and both had more than 45% bounce rates. Although emails accounted for promoting only 1.4% of all sessions, they had a low bounce rate of 17% and a long average session duration of 5 minutes 6 seconds with a conversion rate of 25%, collectively indicating a relatively strong engagement.

Approximately two-thirds or 67.6% (2079/3076) of the users belonged to Canada. Users from Canada also had a relatively low bounce rate (34.35%), high number of pages viewed per session (6.57 pages per session), and long session duration (6:10). However, the users accessed the platform from around the world (Figure 4).

Table 3. Entrance and exit rates for the most viewed pages.

Page	Entrances n (%)[a]	Exits (%)[b]	Bounce rate (%)
Home page	3487 (65.6)	29.2	37.6
Depression in Canada	115 (2.2)	73.4	86.1
Self-Help Exercises	70 (1.3)	14	47.9
Mindsteps	62 (1.2)	5.4	24.2
Screener	55 (1.0)	32.0	50.9

[a]The numbers do not add up to 100% because only several of the most viewed pages are included in the table.

[b]The exit rate is calculated by the number of exits/number of times that page was viewed. Thus, the added percentages are higher than 100% where each row has different number of exits and the number of pages viewed.

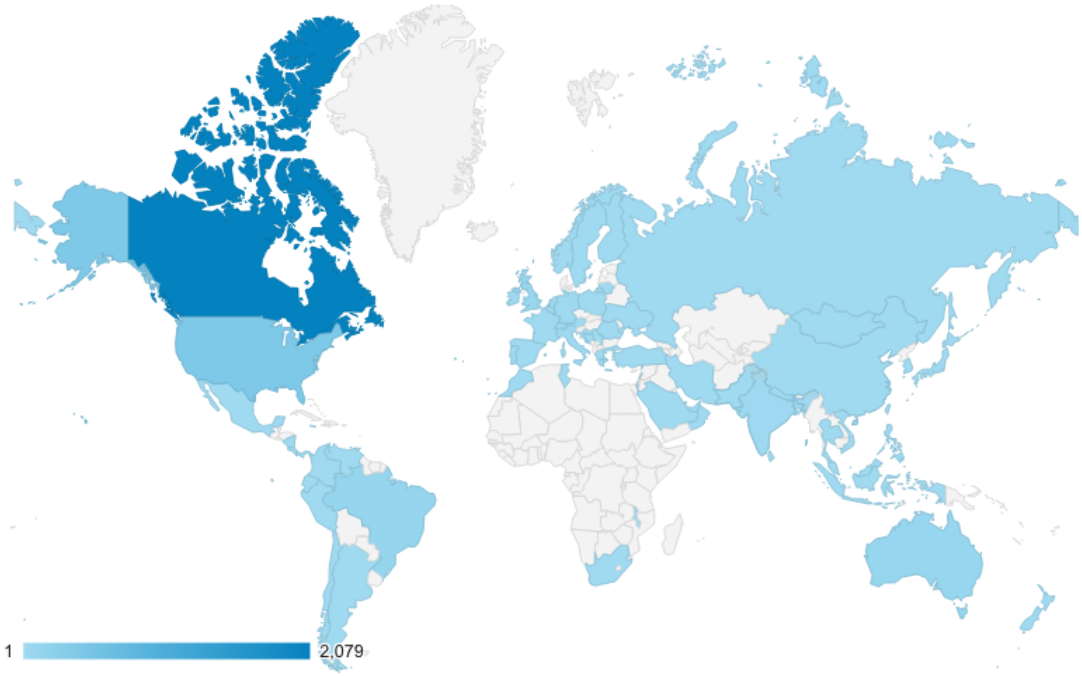
Table 4. Devices used to access WalkAlong.

Device	Sessions (N=5318), n (%)	Bounce rate (%)	Pages per session, n	Mean session duration	Conversion rate (%)
Desktop	4378 (82.32)	39.6	6.17	5 min 43 s	22.7
Mobile phone	677 (12.7)	61.6	2.15	1 min 53 s	7.2
Tablet	263 (5.0)	48.7	3.08	3 min 15 s	11.8

Table 5. Proportion of total sessions for each type of channel.

Channels	Sessions (N=5318), n (%)	Bounce rate (%)	Pages per session, n	Mean session duration	Conversion rate (%)
Direct Traffic	2,420 (45.51)	36.6	7.4	6 min 38 s	24.4
Organic Search	1,256 (23.62)	50.4	3.5	3 min 42 s	11.3
Referrals	849 (16.0)	46.9	3.5	3 min 46 s	22.6
Social Media	717 (13.5)	48.8	4.6	3 min 44 s	18.0
Email	76 (1)	17.1	10.5	7 min 39 s	25.0

Figure 4. Map overlay about locations of users from Google Analytics.



Discussion

Overall Engagement

The first year of operation for the WalkAlong platform saw a total of 3076 users, amounting to 5318 sessions of 5 minutes 6 seconds on average, 29,299 page views, and 31.7% goal completion rate. However, the high proportion of sessions comprising first visits and short session durations suggest a degree of disengagement for the users (Tables 1 and 2). Although the reason for the disengagement is unclear, as it could also be due to acquiring information that is needed early on, this finding does not contradict the pre-existing concerns about lack of engagement evidenced by Web-based mental health interventions [3-5]. These results call for further efforts to continuously improve the platform to be more engaging.

Platform Improvement

The platform can also be improved based on user behavior. For instance, there could be further efforts to improve engagement with tools such as the "Mindsteps" and "Self-Help Exercises." The mean duration of 35 seconds spent on Mindsteps or 1 minute 8 seconds spent on Self-Help Exercises may be deemed too short considering that they are important components of WalkAlong intended to improve mental health outcomes. The relatively short time spent on "Mindsteps" and "Self-Help Exercises" needs to be addressed using better engagement strategies such as improved web design or involving youth to be part of the design process [26]. The WalkAlong home page had the highest entrance rate, indicating that users started their session from this page. This landing information reinforces the role of home page serving as the first impression, and determining the subsequent user behavior [27].

In terms of the devices used to access the WalkAlong platform, the site was viewed mostly via desktops. However, as the usage of mobile phones is ubiquitous among youth, future improvements in WalkAlong may benefit from making the platform more accessible and engaging for mobile phone users [28,29]. A next step could involve developing a native mobile phone app version of the WalkAlong website. This could allow users to access the platform wherever they are without the need of a desktop.

Marketing Strategy

The data indicate that some form of personal referrals indicated by either email or prior knowledge of the URL (direct traffic) results in a relatively stronger engagement (ie, longer average duration, more pages viewed per session, etc) than less personal channels such as referrals through social media or organic searches. In other words, direct referrals such as word-of-mouth strategies among peers could help increase the number of engaged users [30]. This may also include engaging with health care professionals so they can share the platform's URL with their clients.

When observing the location of the users, 67% (2079/3076) users belonged to Canada. This finding aligns with the limited marketing strategy used in Vancouver. WalkAlong can be used by all English-speaking countries, but it can also be used as a template for other platform developments internationally. The WalkAlong team could consider spreading awareness beyond Canada to ensure that such a resource is available to as much youth population as possible.

Using Google Analytics as a Tool for Process Evaluation

Although Google Analytics has provided promising data on the usage patterns of the WalkAlong platform, the tool should be used with careful consideration. For example, comparing the results across various interventions is currently difficult as they serve different purposes with different standards in the number

of users, sessions, and page views [19,25]. For instance, Crutzen et al's website about sexual health showed 850,895 visitors with 5 minutes 6 seconds of average visiting time from March 2009 to December 2010 [19]. It can be assumed that this is a much higher number of visitors with similar duration of visiting time. However, with different periods of time being evaluated for a website serving different purposes, it is difficult to establish the standards for success. Instead, the overall engagement numbers, in particular, may be used to observe trends in usage across different time periods where continual evaluation of the platform is encouraged.

Google Analytics also conforms to a marketing perspective of Web-based behavior rather than to a full evaluation of user behavior [20]. Thus, some variables and information available may not reflect scientific inquiry. This is further complicated by the fact that Google Analytics provides aggregate data where testing of statistical significance for rigorous research purposes can be difficult. Furthermore, the validity of various indicators in measuring user engagement is yet to be established. The number of users may be inaccurate as a new client id is given every time the user deletes the browser cookies, switches devices, or uses a different browser. This may result in the same user being counted as a new user [20]. In addition, long session durations may not, in fact, indicate that users are engaging with the content or a high bounce rate may not indicate that users exit the page due to disinterest quickly, as they could have just quickly found the relevant information they needed. A more detailed analysis of longitudinal user data or a mixed-method assessment to supplement Google Analytics will be important for a more comprehensive process evaluation [31]. For instance, as this study looks only at the first year of operation of WalkAlong, a future study could examine subsequent years comparing internet traffic following changes to the platform, some of which are based on the recommendations mentioned in this paper. Overall, Google Analytics, in combination with other evaluation methods (eg, focus groups, surveys, etc), will provide more accurate interpretations when conducting process evaluation.

Conclusion

Google Analytics was helpful in informing the process evaluation of an open-access Web-based mental health platform. The process evaluation provided information about marketing strategies as well as the aspects of the platform that required improvement. Ideas for future improvements may include marketing the WalkAlong platform outside Canada to get more users from other countries and making the platform more accessible and engaging for mobile users. The rich aggregate data, when combined with other evaluation methods, may provide more accurate interpretations to reinforce or challenge these ideas. Therefore, future studies should focus on developing a mixed methodology that includes Google Analytics to conduct process evaluation of open-access Web-based mental health platforms. With high-quality process evaluation, Web-based mental health interventions may be not only effective but also engaging.

Acknowledgments

Bell Canada Let's talk mental health initiative supported the development of *WalkAlong.ca*

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of pages from WalkAlong.

[PPTX File, 1MB - mental_v5i3e50_app1.pptx]

References

1. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
2. Griffiths KM, Christensen H. Review of randomised controlled trials of Internet interventions for mental disorders and related conditions. Clin Psychol 2006 Mar;10(1):16-29. [doi: 10.1080/13284200500378696]
3. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014 Sep 16;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
4. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
5. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
6. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

7. 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]
8. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
9. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an Internet intervention for depression. J Affect Disord 2018 Dec 15;229:443-449. [doi: 10.1016/j.jad.2017.12.028] [Medline: 29331706]
10. Bruwer B, Sorsdahl K, Harrison J, Stein DJ, Williams D, Seedat S. Barriers to mental health care and predictors of treatment dropout in the South African Stress and Health Study. Psychiatr Serv 2011 Jul;62(7):774-781 [FREE Full text] [doi: 10.1176/ps.62.7.pss6207_0774] [Medline: 21724791]
11. Byrne L, Schoeppe S, Bradshaw J. Recovery without autonomy: Progress forward or more of the same for mental health service users? Int J Ment Health Nurs 2018 Feb 15. [doi: 10.1111/inm.12446] [Medline: 29446509]
12. Davis MJ, Addis ME. Predictors of attrition from behavioral medicine treatments. Ann Behav Med 1999 Dec;21(4):339-349. [doi: 10.1007/BF02895967] [Medline: 10721442]
13. Pampallona S, Bollini P, Tibaldi G, Kupelnick B, Munizza C. Patient adherence in the treatment of depression. Br J Psychiatry 2002 Feb;180:104-109. [Medline: 11823317]
14. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
15. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention. Centers for Disease Control and Prevention. Types of evaluation URL: https://www.cdc.gov/std/Program/pupestd/Types%20of%20Evaluation.pdf [accessed 2018-02-26] [WebCite Cache ID 6xVtKKSE4]
16. Bhattacharya A, Hopkins S, Sallis A, Budd EL, Ashiru-Oredope D. A process evaluation of the UK-wide Antibiotic Guardian campaign: developing engagement on antimicrobial resistance. J Public Health (Oxf) 2017 Jun 01;39(2):e40-e47. [doi: 10.1093/pubmed/fdw059] [Medline: 27451416]
17. Steckler A, Linnan L, editors. Process evaluation for public health interventions and research. San Francisco, CA: Jossey-Bass; Dec 01, 2004.
18. Volker D, Zijlstra-Vlasveld MC, Brouwers EPM, van DFCM. Process Evaluation of a Blended Web-Based Intervention on Return to Work for Sick-Listed Employees with Common Mental Health Problems in the Occupational Health Setting. J Occup Rehabil 2017 Jun;27(2):186-194 [FREE Full text] [doi: 10.1007/s10926-016-9643-4] [Medline: 27150734]
19. Crutzen R, Roosjen JL, Poelman J. Using Google Analytics as a process evaluation method for Internet-delivered interventions: an example on sexual health. Health Promot Int 2013 Mar;28(1):36-42 [FREE Full text] [doi: 10.1093/heapro/das008] [Medline: 22377974]
20. Clark D, Nicholas D, Jamali HR. Evaluating information seeking and use in the changing virtual world: the emerging role of Google Analytics. Learn Pub 2014 Jul 01;27(3):185-194. [doi: 10.1087/20140304]
21. Kirk M, Morgan R, Tonkin E, McDonald K, Skirton H. An objective approach to evaluating an internet-delivered genetics education resource developed for nurses: using Google Analytics™ to monitor global visitor engagement. J Res Nurs 2012 Oct 29;17(6):557-579. [doi: 10.1177/1744987112458669]
22. McCloskey EV, Johansson H, Harvey NC, Compston J, Kanis JA. Access to fracture risk assessment by FRAX and linked National Osteoporosis Guideline Group (NOGG) guidance in the UK-an analysis of anonymous website activity. Osteoporos Int 2017 Dec;28(1):71-76. [doi: 10.1007/s00198-016-3696-2] [Medline: 27438128]
23. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]
24. Makkar SR, Howe M, Williamson A, Gilham F. Impact of tailored blogs and content on usage of Web CIPHER - an online platform to help policymakers better engage with evidence from research. Health Res Policy Syst 2016 Dec 01;14(1):85 [FREE Full text] [doi: 10.1186/s12961-016-0157-5] [Medline: 27905937]
25. Vona P, Wilmoth P, Jaycox LH, McMillen JS, Kataoka SH, Wong M, et al. A web-based platform to support an evidence-based mental health intervention: lessons from the CBITS web site. Psychiatr Serv 2014 Nov 01;65(11):1381-1384 [FREE Full text] [doi: 10.1176/appi.ps.201300512] [Medline: 25124275]
26. Thabrew H, Stasiak K, Merry S. Protocol for Co-Design, Development, and Open Trial of a Prototype Game-Based eHealth Intervention to Treat Anxiety in Young People With Long-Term Physical Conditions. JMIR Res Protoc 2017 Sep 22;6(9):e171 [FREE Full text] [doi: 10.2196/resprot.7250] [Medline: 28939545]
27. Lindgaard G, Fernandes G, Dudek C, Brown J. Attention web designers: You have 50 milliseconds to make a good first impression!. Behav Inform Technol 2006 Mar;25(2):115-126. [doi: 10.1080/01449290500330448]
28. Rainie L, Smith A. Pew Research Center. Washington, D.C; 2013 Oct 18. 35 % of those ages 16 older own tablet computers URL: http://www.pewinternet.org/files/old-media//Files/Reports/2013/PIP_Tablets%20and%20e-readers%20update_101813.pdf [accessed 2017-07-28] [WebCite Cache ID 6sHN1JlyW]

29. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: Integrating smartphone technology in behavioral healthcare. Prof Psychol:Res Pract 2011;42(6):505-512. [doi: 10.1037/a0024485]
30. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries N. Effectiveness of online word of mouth on exposure to an Internet-delivered intervention. Psychol Health 2009 Jul 01;24(6):651-661. [doi: 10.1080/08870440802521094] [Medline: 20205018]
31. Kaushik A. Web analytics: an hour a day. Indianapolis, IN: Wiley Publishing, Inc; 2007.

Abbreviations

PTSD: posttraumatic stress disorder

Edited by J Torous; submitted 07.09.17; peer-reviewed by A Mavragani, F Akbar, T Irizarry, F Solano Zapata; comments to author 01.10.17; revised version received 26.02.18; accepted 19.05.18; published 20.08.18.

Please cite as:
Song MJ, Ward J, Choi F, Nikoo M, Frank A, Shams F, Tabi K, Vigo D, Krausz M
A Process Evaluation of a Web-Based Mental Health Portal (WalkAlong) Using Google Analytics
JMIR Ment Health 2018;5(3):e50
URL: http://mental.jmir.org/2018/3/e50/
doi:10.2196/mental.8594
PMID:30126832

Original Paper

Patient Willingness to Consent to Mobile Phone Data Collection for Mental Health Apps: Structured Questionnaire

Daniel Di Matteo[1], MASc; Alexa Fine[2], BSc (Hons); Kathryn Fotinos[2], BSc (Hons); Jonathan Rose[1], PhD, FRSC; Martin Katzman[2,3,4,5], MD, FRCPC

[1]The Centre for Automation of Medicine, The Edward S Rogers Sr Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

[2]START Clinic for Mood and Anxiety Disorders, Toronto, ON, Canada

[3]Department of Psychology, Lakehead University, Thunder Bay, ON, Canada

[4]The Northern Ontario School of Medicine, Thunder Bay, ON, Canada

[5]Adler Graduate Professional School, Toronto, ON, Canada

Corresponding Author:
Daniel Di Matteo, MASc
The Centre for Automation of Medicine
The Edward S Rogers Sr Department of Electrical and Computer Engineering
University of Toronto
DL Pratt Building
6 King's College Road
Toronto, ON,
Canada
Phone: 1 416 978 6992
Fax: 1 416 946 8734
Email: dandm@ece.utoronto.ca

Abstract

Background: It has become possible to use data from a patient's mobile phone as an adjunct or alternative to the traditional self-report and interview methods of symptom assessment in psychiatry. Mobile data–based assessment is possible because of the large amounts of diverse information available from a modern mobile phone, including geolocation, screen activity, physical motion, and communication activity. This data may offer much more fine-grained insight into mental state than traditional methods, and so we are motivated to pursue research in this direction. However, passive data retrieval could be an unwelcome invasion of privacy, and some may not consent to such observation. It is therefore important to measure patients' willingness to consent to such observation if this approach is to be considered for general use.

Objective: The aim of this study was to measure the ownership rates of mobile phones within the patient population, measure the patient population's willingness to have their mobile phone used as an experimental assessment tool for their mental health disorder, and, finally, to determine how likely patients would be to provide consent for each individual source of mobile phone–collectible data across the variety of potential data sources.

Methods: New patients referred to a tertiary care mood and anxiety disorder clinic from August 2016 to October 2017 completed a survey designed to measure their mobile phone ownership, use, and willingness to install a mental health monitoring app and provide relevant data through the app.

Results: Of the 82 respondents, 70 (85%) reported owning an internet-connected mobile phone. When asked about installing a hypothetical mobile phone app to assess their mental health disorder, 41% (33/80) responded with complete willingness to install with another 43% (34/80) indicating potential willingness to install such an app. Willingness to give permissions for specific types of data varied by data source, with respondents least willing to consent to audio recording and analysis (19% [15/80] willing respondents, 31% [25/80] potentially willing) and most willing to consent to observation of the mobile phone screen being on or off (46% [36/79] willing respondents and 23% [18/79] potentially willing).

Conclusions: The patients surveyed had a high incidence of ownership of internet-connected mobile phones, which suggests some plausibility for the general approach of mental health state inference through mobile phone data. Patients were also relatively

willing to consent to data collection from sources that were less personal but expressed less willingness for the most personal communication and location data.

(JMIR Ment Health 2018;5(3):e56) doi:10.2196/mental.9539

KEYWORDS

passive sensing; mobile phone sensing; psychiatric assessment; mood and anxiety disorders; digital privacy; mobile apps; mobile phone; consent

Introduction

The assessment of mood and anxiety disorders is most commonly performed through clinician interview of the patient or patient self-report. These assessments question patients about their feelings and actions in situations throughout their daily life, asking patients to report the intensity and severity of their symptoms. A review by Meyer et al [1] found that although psychological and medical tests/assessments are valid overall, they have varying degrees of validity, from very low to very high, due to variations in assessment style and responses. Patients may be unwilling or unable to answer accurately due to inaccuracy of recollection [2], psychopathology and/or biases [3,4], ambiguity in the questions, or lack of comfort with their clinician, among other factors. Furthermore, the rating scales employed as part of these assessments are typically not objective—for example, asking the patient to rate their feelings of fear or anxiety, for which no objective measures exist.

An alternative to assessing mental disorders based on descriptions of feelings would be an assessment in which patient behaviors are observed and qualified as indicating increased or decreased severity of the symptoms of the patient's mental disorder. Researchers have begun taking steps toward behavior-based assessments of mental disorders by employing mobile phones as sensing platforms to directly measure or otherwise infer behaviors. Mobile phones are an excellent tool for this purpose, since most owners carry them throughout their day [5] and a combination of sensor data and software techniques can be used to infer high-level behavior and contextual awareness [5-9]. Perhaps most importantly, a mobile phone offers the potential to collect objective data from the patient in contrast to the subjective information provided in interviews or self-reports. The data can be considered objective because there is no human interpretation or mediation of information that comes from sensors that directly measure physical properties. This objectivity of the measurement gives rise to the potential for objective assessment of mental health based on those data [10].

Prior research with this style of passively-sensed, mobile phone–based assessment has shown the ability to measure patients' mental health states relating to disorders such as depression [11] and bipolar disorder [12,13]. Abdullah et al [12] used mobile phone data to assess the social rhythm metric, a marker of stability for individuals with bipolar disorder. They passively recorded geolocation, ambient light levels, communication activity, and ambient audio to infer the metric. Faurholt-Jepsen et al [11] reported that they could classify the affective state of patients with bipolar disorder by analyzing features of the patients' voices as recorded by the mobile phone app developed. Ben-Zeev et al [13] demonstrated that changes in depression severity were associated with changes in physical activity, speech duration, and sleep duration, all of which were measured passively through a mobile phone app. Place et al [15] generated behavioral indicators from mobile phone–collected data that were predictive of clinically assessed symptoms of depression and posttraumatic stress disorder [15].

While these techniques show promise as novel, objective assessment tools, they rely upon the collection and processing of large amounts of private data from a patient's mobile phone. This presents several challenges when requesting patient consent to provide this data. Records of a patient's whereabouts and communications could result in health care providers being subpoenaed for these records by law enforcement agencies. As a result, patients may not be willing to consent to wholesale collection of these data for fear of potential legal ramifications. Patients may also feel uneasy knowing that their health care providers have the ability to scrutinize their actions and communications on a very fine-grained level. This is a specific concern for patients with anxiety disorder. In particular, for those with social anxiety disorder, where a source of anxiety is the fear of judgment from others [16], personal data collected from a mobile phone could potentially form the basis of that judgment. Furthermore, a patient may also fear the possibility that their data could accidentally become public and reveal confidential thoughts, feelings, and behaviors or that perhaps there could be legal implications if the government were to have access to the data [17].

For the researchers and practitioners interested in designing, experimenting with, and deploying these kinds of mobile phone–based assessment tools, it will be important to have a sense of the patient population's willingness to consent to the necessary data collection. A number of studies have surveyed the general consumer population to determine the adoption of Internet of Things [18-20] and wearable technologies [21,22] for health care purposes, and all clearly identified privacy concerns surrounding the data collected by these technologies. However, it is important to know specifically which sources of data available for collection on mobile phones are of most concern and therefore least likely to achieve consent. This question was also raised by Torous et al [23], who, in their study of patient interest in using mobile apps to monitor their mental health conditions, state that they did not address specifically to which sources of information patients would be willing to grant access. This information would allow researchers and developers to build systems that do not rely on unlikely-to-consent sources of data or to do extra work to find ways to address the concerns of patients on particular sources of data collection. It also gives insight into potential barriers that clinicians and researchers

would need to address in order to deploy such systems in a health care setting.

The objectives of this research were as follows: first, since mobile phone–based assessment requires that patients own a mobile phone and use it regularly, it is necessary to measure the ownership rates of mobile phones within the patient population. We also seek to gauge the patient population's willingness to enroll in research studies in which their mobile phone would be used as an experimental assessment tool for their mental health disorder. Finally, and most importantly, we seek to determine how likely patients would be to provide consent for each individual source of mobile phone–collectible data across the wide variety of potential data sources.

Methods

Participants and Procedure

Participants included 82 individuals referred to the START (Stress, Trauma, Anxiety, Rehabilitation, Treatment) Clinic for Mood and Anxiety Disorders, a tertiary care mood and anxiety disorder clinic for the management of their symptoms, located in the city of Toronto, Ontario, Canada. Male and female genders were equally represented with 46% (38/82) males, 46% (38/82) females, and 7% (6/82) individuals who did not respond to a question on gender. The average age of the participants was 41 (SD 14.0) years old. All new intakes and existing patients (ie, any patients in the waiting room) from August 2016 to October 2017 were recruited for the survey while waiting for their appointment with a clinician. Patients were asked to complete a pen-and-paper questionnaire designed to achieve the goals stated above. The questionnaire underwent ethics review by Optimum Clinical Research (protocol number WS2382578), and all respondents provided informed consent before completing the questionnaire.

Materials

The questionnaire designed for this study consisted of 13 self-report questions. The introduction to the questionnaire provides context to the questions which follow by presenting the concept of mobile phone apps as a potential supplement or replacement to questionnaire-based methods of mental health assessment. It is explained that the collection and analysis of data from patients' mobile phones may assist their clinicians in providing better care yet may also impact their privacy. The questions, therefore, are to survey people's willingness to provide sources of information to clinicians and researchers in a scenario where a hypothetical mobile phone app were installed onto their personal mobile phone.

The questions are listed in Textbox 1. Questions 1 through 3 assess general mobile phone use, ownership information, and willingness to install an app that might help with mental health. Questions 4 through 12 ask respondents if they would be willing to share a specific source of data available on their mobile phone. The final question asks which specific brand of phone the respondent uses. For the 2 most potentially rich sources of information, SMS (short message service, or text) messages and raw ambient audio recordings made using the device's microphone, multiple questions are posed in which the amount of information and granularity of the data collection are varied. For example, question 5 asks if respondents would allow collection of SMS metadata (which doesn't contain the content of the message but surrounding information such as who the message was sent to/from and when message occurred), while question 6 asks respondents if they would allow analysis into the contents of their messages for the purposes of extracting potentially clinically relevant words.

Questions 10 through 12 are audio-related. Question 10, concerning the least detailed of the 3 audio data sources, asks respondents if they would share audio metadata, while question 11 concerns willingness to have speech recognition (word detection) performed upon their audio. Question 12 assesses willingness on the most detailed personal data: the unrestrained analysis of ambient audio.

Analysis

Responses to questions 1 through 12 were coded as ordinal variables with 2 (yes and no responses) or 3 levels (yes, maybe, and no responses). When respondents chose to use the other categories of response and provided free-form text, these responses were interpreted as either a yes, maybe, or no response and coded accordingly. Responses to question 13 (on the phone brand type) were coded as a categorical variable. To test for correlation between responses to questions and respondent age, the Spearman rank correlation coefficient (Spearman ρ) was computed along with P values to test against the alternative hypothesis that the correlation was nonzero (using the exact permutation test). To test for associations between responses to questions and respondent gender, responses were cross-tabulated by gender and a chi-square test for independence was performed (2-tailed). All statistics and tests were computed using MATLAB software version 2014b (MathWorks).

Textbox 1. Survey questions.

1. Do you own a mobile phone and use it daily?

(a) Yes, (b) Yes, but not daily, (c) No

2. Do you connect to the internet on your mobile phone, either using a mobile data plan or Wi-Fi?

(a) Yes, (b) No

3. Would you be willing to install and use a mobile phone app to help your doctor better diagnose mental health problems and/or provide treatment?

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

4. Would you be willing to have the app collect and share your location? This would use your mobile phone's Global Positioning System and would pinpoint your location on a map from time to time throughout the day.

(a) Yes (b) Maybe, but I would need to know more information first (c) No (d) Other—specify

5. Would you be willing to have the app record the number of contacts you call or send SMS (text messages) to and the dates and times when you phone or text them? The identities of your contacts would not be shared.

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

6. Would you be willing to have the app read the contents of your text messages to look for keywords related to mental health? (For example, looking for the use of words like "tired," "depressed," or "happy.") The whole text messages would not be shared, only detected keywords.

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

7. Would you be willing to have the app record every time you create a calendar entry in your calendar app? The specifics of the calendar entry or event would not be shared, only the date and time that you create or modify it.

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

8. Would you be willing to have the app record every time you turn your phone's screen on or off (using the power or lock button)?

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

9. Would you be willing to have the app use its sensors to try and detect if and when you are walking, running, in a car, or standing still?

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

10. Would you be willing to have the app occasionally turn on and use the phone's microphone to record the sounds of your surroundings? This audio would be used by the app to try and classify your surrounding as loud, quiet, or busy, but it would not attempt to recognize any words that you or people around you speak nor would it be listened to by humans. It would not record your phone calls, only ambient audio from time to time when you are not making a phone call.

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

11. Consider the same scenario as question 10, but in this case the app will also detect and recognize specific words being spoken aloud by you or anyone else present in the recording. The app would attempt to recognize specific keywords like "tired," "depressed," or "happy." This speech recognition would be done in software by the app, and the audio will never be listened to by humans.

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

12. Consider the same scenario as question 11, but in this case the app is also able to perform any software-based processing to the recorded audio in order detect things that may be relevant to your mental health. This processing (in whatever form) would be done in software by the app, and the audio recordings will never be listened to by humans, nor will humans ever read a transcript of the recordings.

(a) Yes, (b) Maybe, but I would need to know more information first, (c) No, (d) Other—specify

13. What type of phone do you use daily? If you use multiple phones (work and personal), please select whichever corresponds to your personal phone.

(a) iPhone, (b) Android, (c) Blackberry, (d) Windows Mobile phone, (e) Other—specify

Results

Of the 82 respondents, 73 (89%) reporting owning a mobile phone and using it daily (question 1). Rates of internet usage are also high, with 85% (70/82) of respondents reporting that they connect to the internet using their mobile phone (question 2). All respondents reported owning either iPhone, Android, Blackberry, or Windows Mobile mobile phones; Apple iPhones constituted 45% (35/78) of all mobile phones, Android devices constituted 57% (37/78), Blackberry constituted 6% (5/78), and Windows Mobile the remaining 1% (1/78) (question 13). Regarding question 3, willingness to install and use a mobile phone app to help their doctor better diagnose mental health problems and/or provide treatment, 41% of respondents (33/80) indicated that they would be willing to install and use such an app, with another 43% of respondents (34/80) indicating that they may be willing to use such an app, provided they were given more information about how the app functioned. Only 16% respondents (13/80) indicated absolute unwillingness to use such an app. Table 1 presents how responses to these questions (questions 1, 2, 3, and 13) correlate with respondent age and how the responses are associated with respondent gender.

Survey questions 4 through 12 asked respondents if they would be willing to grant the hypothetical mental health monitoring app permission to collect a variety of data sources from their mobile phone. The responses corresponding to each permission (data source) are presented in Table 2. Figure 1 provides a graphical representation of this data, along with respondent willingness to install the app. It is worth noting that nearly none of the survey questions were strongly correlated with respondent age or associated with respondent gender. There is a weak positive correlation, however, between willingness to consent to audio recording and respondent age.

Table 1. Correlation with respondent age and association with respondent gender for mobile phone statistics and general willingness to install a mental health monitoring app.

Question topic	Correlation with respondent age		Association with respondent gender	
	ρ	*P* value	X^2_2	*P* value
Mobile phone ownership (Q1)	0.08	.50	2.81	.24
Internet-connected mobile phone usage (Q2)	–0.09	.46	1.58	.21
Mobile phone types owned (Q13)	–0.10	.43	1.52	.68
Respondent willingness to install a mental health monitoring app (Q3)	0.10	.44	1.86	.39

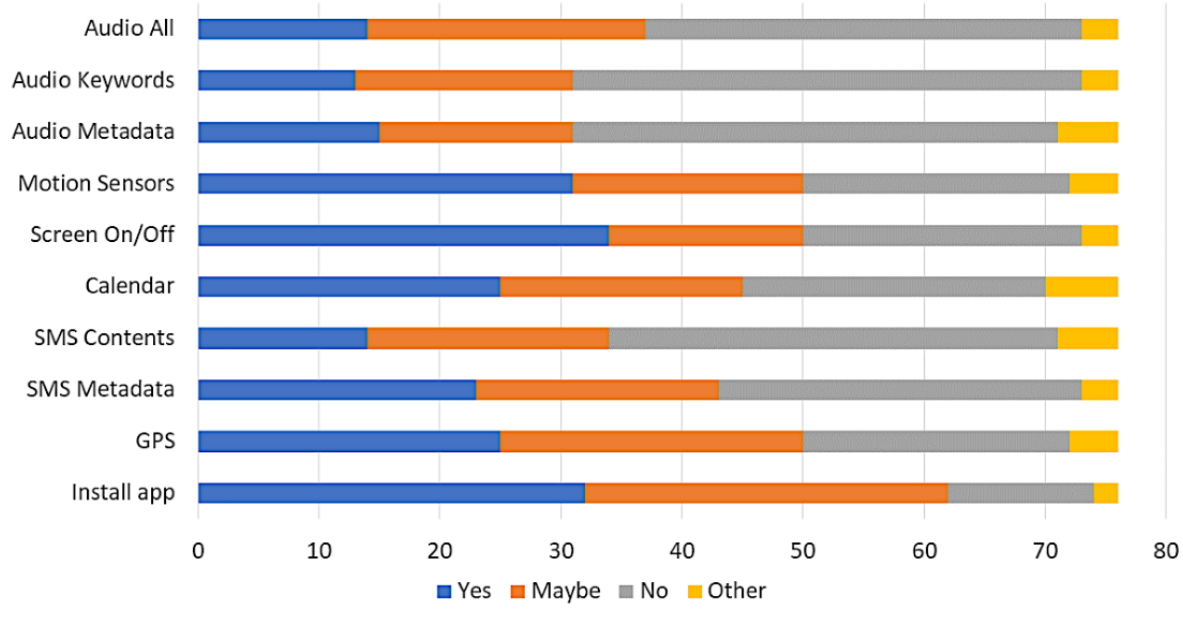
Table 2. Survey respondent willingness to grant permission by category.

Permission	Willing to grant permission, n (%)			Correlation with respondent age		Association with respondent gender	
	Yes	Maybe	No	ρ	*P* value	X^2_2	*P* value
GPS[a] location (Q4)	28 (35)	26 (33)	26 (33)	–0.07	.60	0.18	.91
SMS[b] metadata (Q5)	24 (30)	22 (28)	34 (43)	0.13	.31	1.00	.61
SMS contents (Q6)	16 (20)	21 (27)	42 (53)	0.02	.13	6.30	.04
Calendar (Q7)	26 (33)	26 (33)	27 (34)	–0.03	.83	4.73	.09
Screen on/off (Q8)	36 (46)	18 (23)	25 (32)	0.00	.97	0.48	.79
Motion sensors (Q9)	33 (42)	20 (26)	25 (32)	–0.04	.76	1.18	.55
Audio metadata (Q10)	16 (20)	18 (23)	46 (58)	0.25	.04	4.58	.10
Audio keywords (Q11)	14 (18)	20 (25)	46 (58)	0.29	.02	2.25	.33
Audio unrestrained (Q12)	15 (19)	25 (31)	40 (50)	0.32	.01	4.67	.10

[a]GPS: Global Positioning System.

[b]SMS: short message service.

Figure 1. Respondent willingness to install app and grant permissions. GPS: Global Positioning System; SMS: short message service.



Discussion

Principal Findings

Mobile phone ownership rates in the patient population surveyed in this study are high, with 89% of respondents reporting that they own a mobile phone and use it daily and 85% of patients connect to the internet with it. These results indicate that, in general, lack of mobile phone ownership itself will not present a barrier to the use of mental health monitoring apps in the future. This is further evidenced by American [24] and Canadian [25] mobile phone ownership statistics.

The vast majority of respondents reported owning either an iPhone (45%) or an Android device (47%). This corroborates surveys conducted by Gartner in 2018 [26] which indicate that Android and iOS, collectively, comprise 99.8% of the mobile phone operating systems. As each platform can require significant effort to support, this is relatively good news that suggests that development solely on iOS and Android platforms is sufficient to support the vast majority of patients.

We observe that overall respondents' reception toward the notion of installing a mental health monitoring app is generally positive, with 84% answering either yes or maybe when asked if they would install and use such an app on their mobile phone. This finding shows that there is a general positive interest in using mobile phone apps to aid in mental health assessment. Considering that the questionnaire expresses that such an app is believed to help both clinicians and patients but makes no claims about proven effectiveness nor does it quantify said effectiveness in any way, it is plausible that if such apps are to be developed and proved to be effective that rates of adoption may be higher than reported in this study once patients are presented with these findings.

Respondent willingness to provide access to particular sources of data varies from a minimum of 43% answering yes or maybe in the case of audio metadata or keyword extractions (questions 10 and 11) to a maximum of 68% answering yes or maybe in the case of screen state (question 8). We feel that these results are encouraging considering the survey does not allay any potential fears that may be held by the respondents with regard to data security or data access. In a production-ready app, data security is likely to be a considerable point of focus, with efforts to encrypt data in transit and at rest. Furthermore, it may be possible to develop automated software algorithms to process much of the raw data and report higher level statistics of interest to clinicians [27,28], addressing patient fears that others would be scrutinizing them personally by, for example, reading their text messages or listening to their audio recordings. As our survey did not address any of these points, it may be possible that responses would be more positive given these guarantees.

It is worth noting that data sources that are low resolution in terms of providing personal or private information, such as screen state (Q8) and motion sensor data (Q9), are the most likely to be granted. Contrast this with sources of data that offer much more insight into a person's private life, such as the contents of SMS messages (Q6) and unrestrained audio recording and analysis (Q12), which are among the least likely to be granted (only 20% and 19% yes responses, respectively). This could be interpreted as evidence toward the hypothesis that fear of scrutiny is responsible for unwillingness to provide access to data. If that hypothesis was supported, then this suggests that any automated analysis of data that removes humans from direct observation of the source data may be a method to improve patient willingness to provide data access. Further research is required to explore this hypothesis, however.

Comparison With Other Studies

The rates of mobile phone ownership and use measured in this work is roughly in line with previous research in other areas. The results reported in this study are most similar to those measured in the United States, with a high 70% [29] to 80% range [30]. The predominance of Android and iOS devices within the population under study is also in line with current market statistics [26]. The mobile phone ownership and use measured in this study was somewhat lower than the 96% rate measured by Zhang et al [31] in a Chinese population.

While existing work demonstrates that mobile phone and/or wearable-based health assessment tools are being adopted despite concerns around privacy [21,22], the authors are unaware of any studies that have attempted to determine precisely which sources of data cause most concern. In the broader field of mobile phone apps in general (ie, without a focus on mental health care apps), Felt et al [32] found that SMS messages were the data source that was most cause for concern, more so than Global Positioning System (GPS) location. While they did not consider audio recordings, the identification of SMS messages being more invasive of privacy than GPS location is in line with our results. A similarly broad study into mobile phone user perceptions of privacy and security found that older users exhibited more privacy and security concerns [33]. This is in contrast with the population studied in this work, as we have measured a positive relationship between age and willingness to consent to audio recordings (ie, older people are more willing to consent to audio collection).

Limitations

One limitation of the study is the small, concentrated population of respondents. All respondents presumably live within the Toronto area, and it is not clear how these results may generalize to the greater Canadian population or beyond. Another limitation of the study is the survey design. It is clear that respondents, in general, require more information to provide a sense of their willingness to provide access to data, as the proportion of maybe responses averaged across questions 4 through 12 was nearly one-third of respondents (28%). As mentioned earlier, 2 key pieces of information that would help respondents make more informed decisions are the effectiveness of the app in helping to monitor or manage their mental health if the data are provided and the risks involved and the steps being taken to protect the patient's privacy. Conveying both of these pieces of information to prospective app users will be a challenge for health care providers willing to employ mobile mental health apps.

Another fundamental limitation of this study is one inherent to surveys in general: it is not clear that survey respondents who expressed interest in the hypothetical app and willingness to

consent to collection of their data through the app would actually consent in a real-life scenario involving real mental health apps. We surmise this would depend upon how effective the apps were shown to be and what patient perception of the risks would be.

Finally, it would be interesting for future work to determine if there were any measurable differences between how patients with different disorders might consent to access to their data. One interesting hypothesis to test would be whether patients with social anxiety disorder, who fear evaluation and scrutiny, would be less likely to provide data access, possibly due to the fear of observation or judgment from others characteristic of the disorder [34].

Conclusion

Given the potential for mobile technology to help patients monitor their mental health symptoms in a passive and pervasive way, it is helpful for researchers to understand how patients may respond to requests for access to their personal data. General interest in such an app is moderate, with 41% of respondents indicating they would install a monitoring app and 43% of respondents indicating they may install such an app. Willingness to provide data collection across different sources ranges from 18% to 46%, with more intrusive or private sources of data being more likely to be withheld. Finally, we support previous findings [24,25] that show mobile phone technology adoption alone will not pose a significant problem to fielding mental health apps, as 85% of respondents reported using an internet-connected mobile phone daily.

Acknowledgments

This research was funded in part by the Natural Sciences and Engineering Research Council Discovery Grant number RGPIN-2014-05032.

Conflicts of Interest

MAK has been a consultant or advisory board member for GlaxoSmithKline, Lundbeck, Eli Lilly, Boehringer Ingelheim, Organon, AstraZeneca, Jannsen-Ortho, Solvay, Bristol-Myers Squibb, Shire, Sunovion, Pfizer, Purdue, Merck, Astellas, and Bedrocan. He has undertaken research for GlaxoSmithKline, Lundbeck, Eli Lilly, Organon, AstraZeneca, Jannsen-Ortho, Solvay, Genuine Health, Shire, Bristol-Myers Squibb, Takeda, Pfizer, Hoffman La Rosche, Biotics, Purdue, Astellas, and Forest. He has received honoraria from GlaxoSmithKline, Lundbeck, Eli Lilly, Boehringer Ingelheim, Organon, AstraZeneca, Janssen-Ortho, Solvay, Bristol-Myers Squibb, Shire, Sunovion, Pfizer, Purdue, Merck, Astellas, and Bedrocan. MAK has received research grants from the Canadian Institutes of Health Research, Sick Kids Foundation, Centre for Addiction and Mental Health Foundation, Canadian Psychiatric Research Foundation, Canadian Foundation for Innovation, and the Lotte and John Hecht Memorial Foundation.

References

1. Meyer G, Finn S, Eyde L, Kay G, Moreland K, Dies R, et al. Psychological testing and psychological assessment: a review of evidence and issues. Am Psychol 2001 Feb;56(2):128-165. [Medline: 11279806]
2. Rogler LH, Malgady RG, Tryon WW. Evaluation of mental health. Issues of memory in the Diagnostic Interview Schedule. J Nerv Ment Dis 1992 Apr;180(4):215-222. [Medline: 1556561]
3. van de Mortel T. Faking it: social desirability response bias in self-report research. Aust J Adv Nurs 2008;25(4):40 [FREE Full text]
4. John OP, Robins RW. Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. J Pers Soc Psychol 1994 Jan;66(1):206-219. [Medline: 8126650]
5. Rainie L, Zickuhr K. Americans' views on mobile etiquette, 2015. Washington: Pew Internet and American Life Project URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/08/2015-08-26_mobile-etiquette_FINAL.pdf [accessed 2016-08-31] [WebCite Cache ID 6kAwHhlIt]
6. Harari GM, Müller SR, Aung MS, Rentfrow PJ. Smartphone sensing methods for studying behavior in everyday life. Curr Opin Behav Sci 2017 Dec;18:83-90. [doi: 10.1016/j.cobeha.2017.07.018]
7. Mafrur R, Nugraha IGD, Choi D. Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose. Hum Cent Comput Inf Sci 2015 Oct 12;5(1):31. [doi: 10.1186/s13673-015-0049-7]
8. Pei L, Guinness R, Chen R, Liu J, Kuusniemi H, Chen Y, et al. Human behavior cognition using smartphone sensors. Sensors (Basel) 2013;13(2):1402-1424 [FREE Full text] [doi: 10.3390/s130201402] [Medline: 23348030]
9. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front ICT 2015 Apr 20;2. [doi: 10.3389/fict.2015.00006]
10. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
11. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016 Jul 19;6:e856. [doi: 10.1038/tp.2016.123] [Medline: 27434490]

12. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May;23(3):538-543. [doi: 10.1093/jamia/ocv200] [Medline: 26977102]
13. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
14. Moskowitz DS, Young SN. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. J Psychiatry Neurosci 2006 Jan;31(1):13-20 [FREE Full text] [Medline: 16496031]
15. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. J Med Internet Res 2017 Mar 16;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
16. Clark D, Wells A. A cognitive model of social phobia. In: Heimberg RG, Liebowitz MR, Hope DA, Schneier FR, editors. Social Phobia: Diagnosis, Assessment, and Treatment. 1st edition. New York: Guilford Press; 1995:69-93.
17. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
18. Sicari S, Rizzardi A, Grieco L, Coen-Porisini A. Security, privacy and trust in Internet of Things: the road ahead. Comput Netw 2015 Jan;76:146-164. [doi: 10.1016/j.comnet.2014.11.008]
19. Weber RH. Internet of things: privacy issues revisited. Comput Law Secur Rev 2015 Oct;31(5):618-627. [doi: 10.1016/j.clsr.2015.07.002]
20. Weinberg BD, Milne GR, Andonova YG, Hajjat FM. Internet of Things: convenience vs. privacy and secrecy. Bus Horiz 2015 Nov;58(6):615-624. [doi: 10.1016/j.bushor.2015.06.005]
21. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Ind Manag Data Syst 2015 Oct 19;115(9):1704-1723. [doi: 10.1108/IMDS-03-2015-0087]
22. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: an empirical study from privacy calculus perspective. Int J Med Inform 2016 Apr;88:8-17. [doi: 10.1016/j.ijmedinf.2015.12.010] [Medline: 26878757]
23. 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]
24. Number of smartphone users in the United States from to 2010 to 2022.: Statista; 2015. URL: https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/ [accessed 2018-08-15] [WebCite Cache ID 71h5Ypcuf]
25. Statistics Canada. 2016. Life in the fast lane: how are Canadians managing URL: https://www150.statcan.gc.ca/n1/en/daily-quotidien/171114/dq171114a-eng.pdf [accessed 2018-08-15] [WebCite Cache ID 71h5wepMX]
26. Statistica. Global mobile OS market share in sales to end users from 1st quarter 2009 to 2nd quarter 2017 URL: https://www.statista.com/statistics/266219/global-smartphone-sales-since-1st-quarter-2009-by-operating-system/ [accessed 2018-08-15] [WebCite Cache ID 71h69CKT6]
27. McGraw I, Prabhavalkar R, Alvarez R, Arenas M, Rao K, Rybach D, et al. Personalized speech recognition on mobile devices. ICASSP 2016;1(1):5955-5959. [doi: 10.1109/ICASSP.2016.7472820]
28. Andriotis P, Takasu A, Tryfonas T. Smartphone message sentiment analysis. In: Peterson G, Shenoi S, editors. Advances in Digital Forensics. Berlin: Springer; 2014.
29. Rubanovich CK, Mohr DC, Schueller SM. Health app use among individuals with symptoms of depression and anxiety: a survey study with thematic coding. JMIR Ment Health 2017 Jun 23;4(2):e22 [FREE Full text] [doi: 10.2196/mental.7603] [Medline: 28645891]
30. 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]
31. Zhang H, Zhang H, Wang X, Yang Z, Zhao Y. Analysis of requirements for developing an mHealth-based health management platform. JMIR Mhealth Uhealth 2017 Aug 03;5(8):e117 [FREE Full text] [doi: 10.2196/mhealth.5890] [Medline: 28778840]
32. Felt A, Egelman S, Wagner D. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. SPSM 2012;1(1):33-44. [doi: 10.1145/2381934.2381943]
33. Fife E, Orjuela J. The privacy calculus: mobile apps and user perceptions of privacy and security. Int J Eng Bus Manag 2012 Jan;4:11. [doi: 10.5772/51645]
34. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Arlington: American Psychiatric Publishing; 2013.

Abbreviations

GPS: Global Positioning System
SMS: short message service
START: Stress, Trauma, Anxiety, Rehabilitation, Treatment

Edited by G Eysenbach; submitted 18.12.17; peer-reviewed by S Schueller, J Andrews; comments to author 16.03.18; revised version received 10.05.18; accepted 21.06.18; published 29.08.18.

Please cite as:
Di Matteo D, Fine A, Fotinos K, Rose J, Katzman M
Patient Willingness to Consent to Mobile Phone Data Collection for Mental Health Apps: Structured Questionnaire
JMIR Ment Health 2018;5(3):e56
URL: http://mental.jmir.org/2018/3/e56/
doi:10.2196/mental.9539
PMID:30158102

Original Paper

Recognition of Emotions Conveyed by Touch Through Force-Sensitive Screens: Observational Study of Humans and Machine Learning Techniques

Alicia Heraz[1], MEng, PhD; Manfred Clynes[2], DSc

[1]The Brain Mining Lab, Montréal, QC, Canada

[2]Institute of Amity and Emotion Research, Nyack, NY, United States

Corresponding Author:
Alicia Heraz, MEng, PhD
The Brain Mining Lab
300 Rue Saint-Georges
Montréal, QC, J4P 3P9
Canada
Phone: 1 5145508902
Email: alicia@bmlabca.com

Abstract

Background: Emotions affect our mental health: they influence our perception, alter our physical strength, and interfere with our reason. Emotions modulate our face, voice, and movements. When emotions are expressed through the voice or face, they are difficult to measure because cameras and microphones are not often used in real life in the same laboratory conditions where emotion detection algorithms perform well. With the increasing use of smartphones, the fact that we touch our phones, on average, thousands of times a day, and that emotions modulate our movements, we have an opportunity to explore emotional patterns in passive expressive touches and detect emotions, enabling us to empower smartphone apps with emotional intelligence.

Objective: In this study, we asked 2 questions. (1) As emotions modulate our finger movements, will humans be able to recognize emotions by only looking at passive expressive touches? (2) Can we teach machines how to accurately recognize emotions from passive expressive touches?

Methods: We were interested in 8 emotions: anger, awe, desire, fear, hate, grief, laughter, love (and no emotion). We conducted 2 experiments with 2 groups of participants: good imagers and emotionally aware participants formed group A, with the remainder forming group B. In the first experiment, we video recorded, for a few seconds, the expressive touches of group A, and we asked group B to guess the emotion of every expressive touch. In the second experiment, we trained group A to express every emotion on a force-sensitive smartphone. We then collected hundreds of thousands of their touches, and applied feature selection and machine learning techniques to detect emotions from the coordinates of participant' finger touches, amount of force, and skin area, all as functions of time.

Results: We recruited 117 volunteers: 15 were good imagers and emotionally aware (group A); the other 102 participants formed group B. In the first experiment, group B was able to successfully recognize all emotions (and no emotion) with a high 83.8% (769/918) accuracy: 49.0% (50/102) of them were 100% (450/450) correct and 25.5% (26/102) were 77.8% (182/234) correct. In the second experiment, we achieved a high 91.11% (2110/2316) classification accuracy in detecting all emotions (and no emotion) from 9 spatiotemporal features of group A touches.

Conclusions: Emotions modulate our touches on force-sensitive screens, and humans have a natural ability to recognize other people's emotions by watching prerecorded videos of their expressive touches. Machines can learn the same emotion recognition ability and do better than humans if they are allowed to continue learning on new data. It is possible to enable force-sensitive screens to recognize users' emotions and share this emotional insight with users, increasing users' emotional awareness and allowing researchers to design better technologies for well-being.

(JMIR Ment Health 2018;5(3):e10104) doi:10.2196/10104

KEYWORDS

emotional artificial intelligence; human-computer interaction; smartphone; force-sensitive screens; mental health; positive computing; artificial intelligence; emotions; emotional intelligence

Introduction

Background

Emotions are distinct natural entities that involve the mind and the body. They modulate our physiology by increasing or decreasing variables such as heart rate, respiration, and body temperature, and our psychology by altering perception, beliefs, and virtual images [1,2]. Emotions seek expression, and they use one or many motor outputs (sequentially or simultaneously) to fulfill their need to be expressed: face, voice, arms, and legs are used in a combination of one or many outputs to express emotions depending on the context, the physical condition of the body, the available means of communication, and the choice of the expresser [1].

Touch is a profound form of communicating emotions, and many researchers have studied its use in and impact on health and well-being [1,3-9]. A few minutes of daily touches not only enhance growth and weight gain in children, but also lead to emotional, physical, and cognitive improvements in adults [6,7]. Touch releases hormones and neuropeptides, and stimulates our bodies to react in very specific ways: the levels of blood pressure, heart rate, and cortisol change, and the hippocampus area of the brain is activated for memory [3]. Humans can easily communicate and sense emotions conveyed through touch [8,9]; babies respond well to touch [4] and loving touches are critical to the health of premature infants [5].

We use touch to communicate with many devices in our daily lives. As machine interfaces are engaging users more frequently and tend to mimic human-human interactions to facilitate natural communications, it is becoming key to develop new algorithms for the new interactive and sensitive touch screens to capture and recognize users' emotions and increase the level of emotional intelligence of both users and their devices.

Emotional intelligence, or our ability to recognize and regulate our own emotions and those of others, is key to communicating well. Recognizing emotions when they are expressed and regulating them in ourselves and others helps us achieve effective communications and maintain good mental health [10-12].

Emotional artificial intelligence, or a machine's ability to express emotions and recognize and regulate users' emotions, has also become a key capability of an intelligent machine enabling effective interactions with its users [13]. Various methods are being used and technologies are being built to detect users' emotions as they interact passively or actively with machines. The most commonly used techniques for emotion recognition are one or a combination of many of the following approaches: text, facial expressions, voice tones, biosensors and body movements, and gestures.

Many technologies have implemented facial coding and voice analysis theories to recognize emotions, but none have explored the touch theory [1]. Software technologies analyze the face, tone of voice, and textual natural language to recognize emotions [14]. It is difficult to fulfill the requirements of these technologies and measure emotions accurately in real-life situations. It is essential to find and add other more accessible ways to measure emotions: that is, mobile phone use and touch screen behavior.

Emotion recognition techniques using text process words and sentences in a particular language. The most common techniques process natural language and extract emotions and sentiments from writings and conversations found in books, blogs, chat rooms, and social media platforms [15]. One of the biggest challenges of this method is to recognize emotions in the context of the text. An emotion can be expressed without using the word that denotes it or any of its synonyms. New words and expressions can be created, or words from other languages or sarcasm can be used to express emotions, which makes the task of recognizing emotions very difficult [16].

Usually, facial emotion recognition techniques segment images of the face into specific regions and analyze their movement. Regions of interest include cheek, chin, wrinkles, eyes, eyebrows, and mouth [17]. Different classification techniques are then applied to recognize emotions [18]. New studies have claimed that faces alone do not universally communicate emotions, and that conceptual knowledge supported by language is necessary to distinguish and recognize emotions [19,20]. In addition, a static image does not convey the information related to the dynamic form of the expression to accurately recognize emotions.

Speech emotion recognition techniques analyze the tone of voice and other speech features to detect emotions and their dimensions. Various methods and interfaces have been designed to extract various features from speech signals, and they are usually adapted to a particular language. The tone of speech varies with different cultures. A person from a certain area, talking in a normal tone, might sound angry to someone from another culture due to differences in normal speed and volume between the two cultures. Someone talking in a low and slow tone might appear as sad for some and polite for others. Additionally, speaking in real-life conditions is corrupted with various noises. This makes it hard for a machine to isolate a particular voice and recognize emotions [21].

Emotion recognition using biosensors monitors the physiological variables of the autonomic nervous system (ANS) that are affected by emotions. Biosensors can be invasive or noninvasive and collect variables in the ANS including heart rate, skin conductance, heart rhythm, blood volume, and temperature to recognize emotions through changes in their patterns. Spatial and temporal analysis of the brain's activity are two other techniques that are used to recognize emotion [22]. Functional magnetic resonance imaging focusses on identifying the regions of the brain involved in expressing emotions, and electroencephalography monitors the electrical activity of the brain to recognize emotions. Different clustering and evaluating

techniques are then used on these physiological changes to detect emotions. But consistent and universal patterns in the ANS in relation to emotions have not been found, and many technical challenges are still to be solved [12,23,24].

Body movements and finger gestures are other ways of expressing emotions. Features such as amplitude, speed, fluidity, shape of movements, and motion direction are being extracted from expressive children and adults, and various methods and techniques have been applied to recognize emotions [25-27]. A body action and posture coding system has been developed recently to enhance the understanding of the role of body movements in expressing emotions [28]. But emotion recognition based on body movements and gestures is the least popular way of evaluating emotions. Tracking body movements and gestures in 3 dimensions is difficult and requires many sensors, which is one of the major drawbacks of this method.

Multimodal approaches have also been used to recognize emotions. By analyzing two or many measures from the face, voice, text, or ANS, these approaches provide better accuracy than do individual modalities but are complex and not easy to replicate or scale [29-31].

In the last few years, researchers have started to explore correlations between emotions, features from smartphones, and gestures [28,32]. Analyzing data from a smartphone's accelerometer predicted the emotional dimension of arousal with an accuracy of 75% [33]. Analyzing the length, time, velocity, and pressure of finger strokes on a smartphone predicted the emotional dimension of valence with 84.9% accuracy [34]. Analyzing features extracted from textual contents and user typing predicted anger, disgust, happiness, sadness, neutrality, surprise, and fear with 72% accuracy [35].

The worldwide number of mobile phone users is expected to pass the 5 billion mark by 2019 [36]. Most of the mobile market growth can be attributed to the increasing popularity of smartphones. Users touch their smartphones thousands of times a day [37]: they play games, purchase products, and interact with other users in chat rooms and social media platforms. Users spend 54 minutes to 3.8 hours per day on their smartphone. Among 16- to 24-year-olds, 94% possess a smartphone and spend up to 4 hours per day on it. They open an app every 15 minutes because they feel the urge to do so: it is difficult for most of them to reduce the time spent on a smartphone or control the frequency of its use [38]. Addictive behaviors are dictated by uncontrolled emotions, where reasoning and logical thinking is not applied [2,39]. The World Health Organization classifies addictive behaviors for technology as a mental health problem [40].

It is thus essential to develop technologies for emotional awareness and help users establish a healthy communication between reason and emotions to lower addictive behaviors and suffering, improve decision making, and increase well-being [41].

Objective

This study explored the power of touch and its potential to convey distinct emotions when it is used as a means to communicate with apps and distant users via force-sensitive smartphones. Our ultimate goal is to increase users' emotional awareness by powering smartphone apps with touchscreen emotional intelligence. We hope to open new opportunities for designers to create new interactive emotional experiences, provide emotional inputs for developers to enrich their apps, and offer insight for researchers to better understand emotions in the context of human-computer interactions.

We conducted 2 experiments in this study to recognize anger, awe, desire, fear, hate, grief, laughter, love, and no emotion. Experiment 1 examined whether humans are able to recognize the emotions by looking at passive expressive touches. Experiment 2 collected features of expressive touches of emotionally aware participants with good imagery ability and used machine learning techniques to predict emotions.

Methods

Emotions

Emotions are complex entities that often cannot be defined with just one single word. Naming an emotion is labelling the qualities of its psychophysiological manifestation, and these qualities are not precisely known. Naming by language the experience of anger is not a guarantee of the existence of a simple and clear psychophysiological pattern. Anger is not a simple entity, and complex and mixed patterns may have simple names. Sometimes the naming is, to a degree, confused and confusing. Some emotions remain nameless.

In this study, we were interested in what we define as *biological emotions* or nonverbal emotions for which language is not required when they are communicated. We listed 8 emotions (and no emotion) that we considered to be biological, and we hypothesized that they are easily recognizable by a perceiver if expressed authentically: anger, awe, desire, fear, grief, hate, laughter, and love, as well as no emotion. The main word we coined for each emotion is approximate and not unique. Translating our words to other languages and explaining them to our participants required that we define every individual word by describing how people react and what they say and do when they are under the influence of that emotion. Table 1 describes our 8 biological emotions.

In addition to these descriptions, we added a set of images and videos showing people's faces, speech, and body movements for every emotion. We collected this multimedia content from the internet. We were inspired by the International Affective Picture System [42] and the affective videos of the Laboratoire d'Informatique en Image et Systèmes d'information Annotated Creative Commons Emotional Database [43], and standardized in terms of its audio and visual characteristics (brightness, loudness, distance, color, and size).

Participant Recruitment

We recruited and screened volunteers and smartphone users before assigning them into 2 groups to participate in 2 experiments. We contacted a local volunteer center and posted advertisements on websites asking people above 16 years old who were interested in emotions and technology to participate in our study for free.

Table 1. Labels, descriptions, and synonyms of the 8 biological emotions (and no emotion).

Label	Description	Synonyms
Anger	When you are angry, you boil, react, object, yell, or swear. You say words or expressions like "fuck," "shit," "no," "stop," or other synonyms silently in your head or loudly in your own language, usually your native language.	• Frustration • Rage • Fury
Awe	When you are in awe, you freeze or slow in contemplation. You disconnect from distractions and get absorbed by the object of your awe. You are speechless and cannot link what you discover with what you already know.	• Interest • Discovery • Contemplation
Desire	When you desire, you want, crave, need, and starve for. You say words or interjections like "yummy," "come," "tasty," "want you," or other synonyms silently in your head or loudly in your own language, usually your native language.	• Need • Lust • Want
Fear	When you fear, you withdraw, hide, freeze, or tremble. You remain silent or say words or expressions like "no" or other synonyms silently in your head or loudly in your own language, usually your native language.	• Scare • Panic • Terror
Grief	When you are in grief, you are very sad, and feel helpless and weak. You suffer and feel pain. You cry, moan, and whimper.	• Agony • Mourning • Sadness
Hate	When you hate, you destroy, crush, and break. You say words or expressions like "perish" or "die" in your head or loudly in your own language, usually your native language.	• Detestation • Loathing • Vengefulness
Laughter	When you laugh, your breath and voice are chopped and your eyes twinkle and tear. You repeat "Ha ha" or other sounds while you move in the same rate as you laugh and emit sounds.	• Chuckle • Giggle • Excitement
Love	When you love, you care, protect, comfort, and maintain the state of the loved object. You smile, remain silent, or say words or expressions like "dear," "cute," or "sweet" in your head or loudly in your own language, usually your native language.	• Affection • Delight • Joy
No emotion	When you are not under the influence of an emotion, you reason with ease. Counting from 1 to 10 while seeing or visualizing the numbers in your head is an example of a very simple and unemotional state.	• Reasoning • Thinking • Counting

Grouping Procedure

We invited all volunteers to complete 2 tests: the Levels of Emotional Awareness Scale (LEAS) to assess their emotional awareness [44]; and the Questionnaire Upon Mental Imagery (QMI) to assess their imagery ability [45,46]. For the purpose of our study, and the requirement to express pure and authentic emotions in both experiments, we needed all participants in group A to be not only emotionally aware, but also able to physiologically react to imaginary emotional situations. Only 20 participants among the 117 volunteers accepted to take the tests and apply to be part of group A.

The LEAS is an open-ended test in which we assessed the ability of volunteers to use emotion words in various situations to describe their feelings and the feelings of others [44]. A total of 20 volunteers answered 20 questions describing 20 emotionally evocative situations, and we hand-scored their responses. The scoring was as follow: 0 was given to nonemotional responses or when a response described thought instead of feeling; 1 was given when participants described physical awareness (eg, "I feel tired"); 2 was given when a response described an undifferentiated emotion (eg, "I feel bad") or when the response described an action (eg, "I feel like I'm going to punch him in the face"); 3 was given when feelings were described using discrete emotions (we have a glossary of more than 600 discrete emotions collected from prior studies); 4 was given when many different discrete emotions were used to describe mixed or complex feelings; and 5 was given when participants described and differentiated their own feeling from someone else's feelings using discrete emotions [47].

A higher score in the LEAS correlates positively with empathy, understanding of others, and openness to experience [48], as well as the ability to recognize emotions [49,50].

The QMI is a 600-item measure of mental imagery ability for 7 sensory modalities (visual, auditory, cutaneous, kinesthetic, gustatory, olfactory, and organic). In our study, we focused on the visual, cutaneous, and kinesthetic sensors only (the scoring of each of the 7 modalities is independent, and the emotion induction protocol requires participants to focus on 3 senses). We thus asked our volunteers 80 selected questions similar to the original questions of the test to indicate how clearly they could imagine a series of situations (eg, the form and movement of an assassin approaching the bed, touching silk, running fast to catch a car). Scoring was on a 7-point vividness rating scale varying from 1 ("Perfectly clear and as vivid as the actual experience") to 7 ("I think about it but I cannot imagine it"),

with a total score ranging from 80 to 560. Higher total scores indicate weaker imagery ability. Physiological activity in response to emotional imagery varies as a function of imagery ability. This means that good imagers show greater emotion-specific physiological activity than poor imagers [51].

We conducted experiments 1 and 2 in the laboratory on the same day for group A. Group B participated in experiment 1 remotely. We asked group A participants to follow some instructions (see below) 1 day before the day of the experiments. On the day of the experiments, they participated in an interview session and a training session before experiments 1 and 2 [52].

Instructions for the Day Before

We asked group A to rest, sleep, and eat well but not too much, because emotions are subject to physical and psychological states and can be difficult to induce under conditions of stress, fatigue, lack of sleep, hunger, or heavy meals [1]. For the needs of the training session, we asked them to come, if possible, with a voluntary partner with whom they were able to be emotionally intimate. Alternatively, we set up a time when 2 single group A participants could come together in the same time and partner each other during the training session. We asked all of them to bring their smartphone.

Interview Session

We excluded group A participants if they had a history of any of the following: current alcohol and drug abuse or dependence, neurological disease or trauma, and other medical or psychiatric complications. To share our definition of emotions clearly with everyone in group A, we gave them the list of emotions described in Table 1, and they viewed our collected multimedia content for each emotion. We encouraged them to ask questions and translate the words we gave as examples to their own words and language. All group A participants signed a consent form where a strict ethics code was applied, including their right to ask questions, withdraw from the experiment, keep their personal information private, and understand the scope and the purpose of the study.

Smartphone Test

Our protocol in experiment 2 required a minimal granularity in the measurement of finger pressure and area to collect enough change in data for our machine learning algorithms. To test whether a smartphone is sensitive enough, we designed an app that can be set up on any smartphone to check our requirements. After installing and launching the app on their smartphone, we asked participants to press on their screen as hard as possible for 5 seconds. We required a minimum granularity level of 10 for both area and force, a minimum pixel density of 100 pixels per inch, and a maximum temporal resolution of 1 millisecond. We used the device model Motorola XT1023 (Motorola Mobility LLC, Libertyville, IL, USA) with all of those who did not have a sensitive enough smartphone.

Training Session

The purpose of this session was to train participants from group A to express each emotion as precisely as possible by using only their finger and touching the palm of the partner to communicate each emotion. We called group A participants "expressers" and their partners "perceivers." We encouraged but did not oblige expressers to use their middle finger during the expression because it is the first finger to reach the target of touch and the least cumbersome to use.

We asked the expressers to read the description of each emotion (see Table 1) and use our inductive material (images and videos) to stimulate their imagination and remember an emotion-provoking life situation. We asked them to describe the situation as if they were actively involved and emotional. They were encouraged to close their eyes and imagine the specific situation so as to experience it more accurately. A blank paper was provided, and participants were required to write down a description of the situation, their thoughts, and their words for each emotion. They were encouraged to use their own language. Textbox 1 lists the instructions given to the expressers.

We gave the perceivers, in a separate room, the list of emotions (see Table 1) and asked them to memorize them. They were informed that they were to be blindfolded and seated on a chair with their palm resting on their knee, and that an expresser would try to communicate an emotion from the list by only touching the perceiver's palm with their finger. We asked the perceivers to try to guess the emotion conveyed by the finger and respond with 1 of the following 3 options: 1: "I don't know;" 2: "I hesitate between...;" and 3: "I know; it is..."

We asked expressers to express each emotion once or many times until the perceiver guessed their emotion correctly. They were free to move to another emotion before coming back to an emotion they had already tried to express without being successful and try again until they succeeded. Expressers were given a paper on which the emotions were listed and were also asked to put 1 or 0 at every trial (1: their partner successfully recognized the emotion; 0: their partner was confused or not able to recognize the emotion). Figure 1 shows the emotional expression of participants on the palms of their partners.

We asked expressers who completed the training session to participate in experiments 1 and 2, but only to express the emotions they were able to communicate successfully to their perceivers.

Experiment 1

The objective of this experiment was to answer the following question: if emotions modulate our touches [1,8,9], will humans be able to recognize emotions by only looking at an expressive touch?

Textbox 1. Instructions to induce and express emotions.

Choose an emotion from the list. Remember the related scene that you described as vividly as possible. Imagine as if it were really happening to you. Communicate your emotion by touch as precisely as possible.

Figure 1. Group A participant expressing emotions in the palm of their partner.

We asked participants from group A to express the emotions that their partners were able to recognize during the training session against a transparent glass under which we video recorded their touches for 30 seconds. Each session started with neutral (nonemotional) touches.

We asked participants to express each emotion with one or many successive touches. They were encouraged but not obliged to use their middle finger. Figure 2 shows 3 frames from 3 different videos in which a participant expressed fear, grief, and laughter, respectively (left to right).

We then showed the videos online to group B and asked them to guess the emotion expressed by the finger by classifying the expression as 1 of the 8 emotions (and no emotion) described in Table 1.

Experiment 2

The objective of this experiment was to answer the following question: can we teach machines how to accurately recognize users' emotions from their touches on force-sensitive screens?

We asked group A participants to express the emotions that their partners were able to recognize during the training session against an app that we installed on a smartphone. They were encouraged to use their middle finger because it is the least cumbersome finger and the first to reach and touch the screen, but they were allowed to use any other finger if it was easiest for them to use.

Figure 3 shows the app's interface and a participant expressing an emotion. Participants followed the same instructions as those in the training session (see Textbox 1).

Each session started with neutral (nonemotional) touches. We asked participants to express each emotion with one or many successive touches. They were encouraged to express each emotion at least 10 times, successively or not. After each emotional expression, participants took a 5-minute break to relax and resume a neutral state, thus preventing potential carryover effects of the previous emotional experience. During this period, we asked participants to rate on a single 5-level scale the purity of their emotional expression and the extent to which their finger expression conveyed the emotion. The rating varied from 0 to 5, where 0 was "not expressed" and 5 was "expressed well." This allowed us to assess the accuracy of the induction procedure, as well as the subjective state of the participant after the emotional experience and expression.

Precision, Calibration, and Normalization

We recorded the coordinates (x_t,y_t) of finger touches, their amount of force (F_t), and skin area (S_t), all as functions of time with a resolution of 1 millisecond. Knowing the screen density of each device, we converted (x_t,y_t) from pixels to millimeters with a precision of a hundredth of a millimeter. We calibrated F_t and S_t, knowing the maximum amount of force and skin area a participant could apply on a device, and then normalized it on a scale of 0 to 1 with a precision of 2 digits after the decimal, because the maximum values F_t and S_t differ not only between devices but also between participants. Some devices are more sensitive than others, and the maximum values of F_t and S_t are not always 1. Participants had different finger sizes, and large fingers record larger areas.

Data Filtering, Touches, and Expressions

Among the data we collected, and based on the subjective rating, we retained only the expressions that were rated 3 or more on the scale of purity. We grouped successive touches with a delay of less than 1000 milliseconds and considered them to be part of one single emotional expression. Expressions comprised one or many touches depending on the emotion, its intensity, and the need to express it.

Figure 2. Frames from the expression through touch of fear (left), grief (middle), and laughter (right).

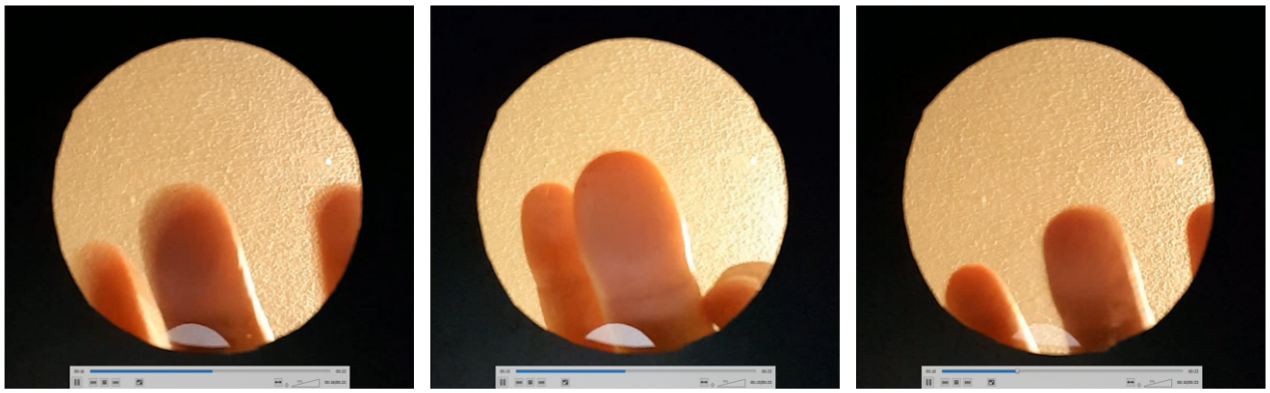


Figure 3. Emotional expressions on the mobile app.

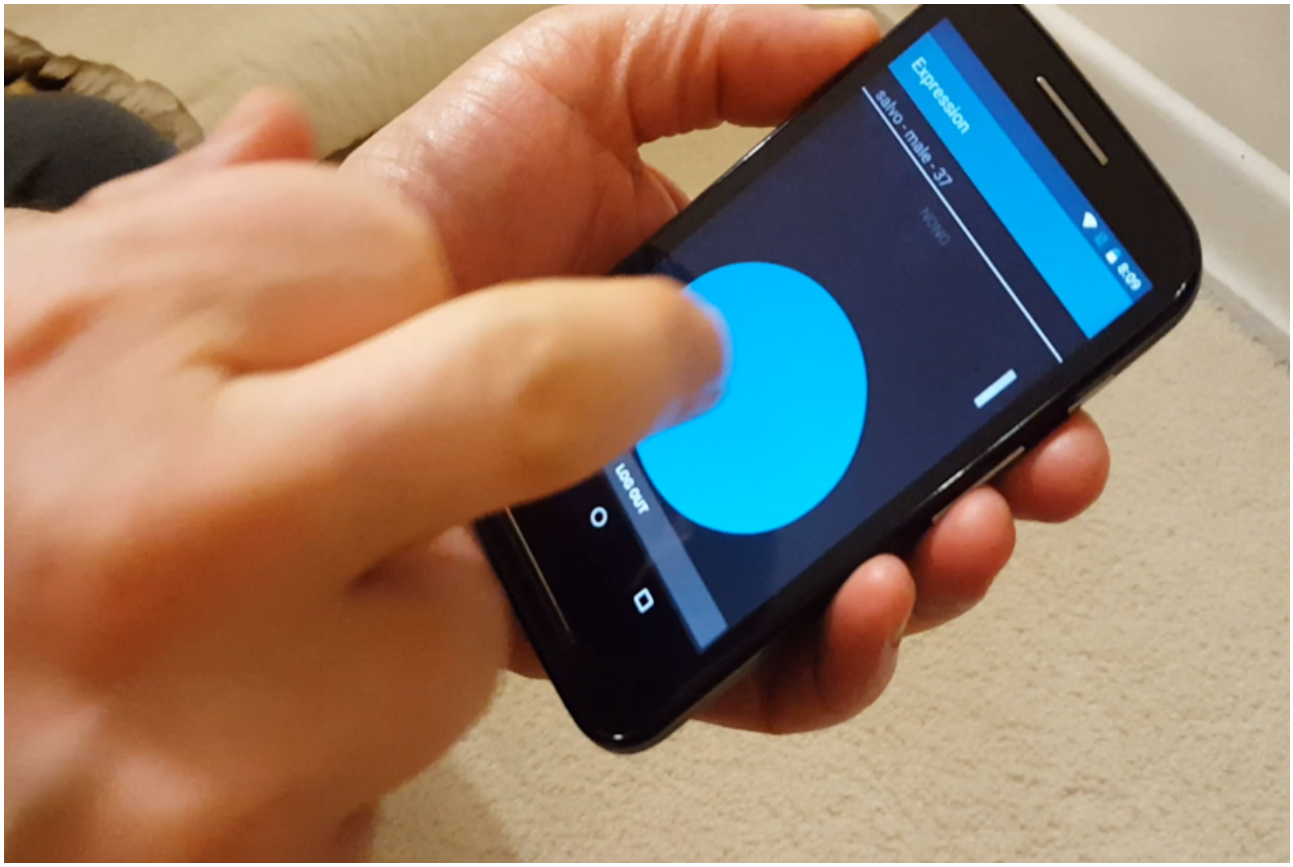


Table 2. Feature dependency test using paired *t* test.

Pair	Mean (SD)	SE	95% CI
1	5.139 (69.361)	1.441	2.313 to 7.966
2	175.118 (1000.759)	20.795	134.339 to 215.897
3	0.756 (1.520)	0.032	0.695 to 0.818
4	–284.59 (3446.514)	71.616	–425.03 to –144.15

Feature Extraction

To set up an initial set of relevant features and observe the expression of each emotion in time, we designed a visualization tool using Python.

Dimension Reduction

Based on our intuitive assumptions and observations and the experimental context, we calculated for every expression an initial set of the following 17 spatiotemporal features: the coordinates (x1, y1, x2, y2) of the beginning and end of an expression; the distance, angle, duration, velocity, and acceleration of the expression; the total number of touches and the mean duration of touches of an expression; and the mean, maximum, and velocity of both force and size of an expression. Further statistical analysis and feature selection techniques including principal component analysis [53] allowed us to reduce the number of features to 9 independent features. We retained the first 9 eigenvectors because they captured 95% of the total variance in the original data. The distribution of variance among the 9 components was 20.1% (roughly corresponds to the duration), 18.2% (the amount of force), 12.7% (the number of touches), 11.3% (the spatial extent on the x,y axis), 10.4% (the angle), 8.2% (the distance), 5.1% (the size), 4.6% (the velocity), and 4.4% (the acceleration).

We used the parametric paired *t* test to measure the degree of variance and dependency between the normally distributed features. Table 2 displays the comparison results.

We see that the mean differences 5.139 (in pair 1), 175.118 (in pair 2), 0.756 (in pair 3), and –284.59 (in pair 4) between features are not equal to zero. With 95% confidence (P=.05), we could conclude that there was a significant statistical score to indicate that our selected features were independent.

Results

Participants

We recruited and screened 117 volunteers and smartphone users between the ages of 16 and 63 years (47 male and 70 female participants). Their mean age was 32.5 (SD 13.9) years.

Only 15 participants were rated as good imagers and highly emotionally aware enough to be part of group A (15/117, 12.8%; 7 male and 8 female participants). They were all right-handed and scored above 70 (very much above average) in the LEAS test and below 107 (above average) in the QMI.

The other 5 participants were either not enough emotionally aware or not good enough imagers, or both. We assigned them to group B, which comprised 102 volunteers (40 male and 62 female participants), and we applied no additional tests or requirements to them. Both groups participated in experiment 1, and only group A participated in experiment 2. Figure 4 shows the distribution of the participants by group and sex.

Smartphone Sensitivity Tests

We tested 11 different smartphone devices: 63% (7/11) were eligible to be used in experiment 2 (see Table 3).

Experiment 1: Human Recognition for Emotions

We collected 102 responses for each emotion. The results proved to be highly successful and significant, with all 8 emotions (and no emotion) correctly recognized with great accuracy with only 1 attempt. Errors of recognition were mainly choosing hate for anger (7/918, 0.8%) and vice versa, or choosing awe for no emotion (11/918, 1.2%) and vice versa. A total of 12 of the 918 participants (1.3%) confused hate with laughter.

Figure 4. Group and sex distributions of the participants.

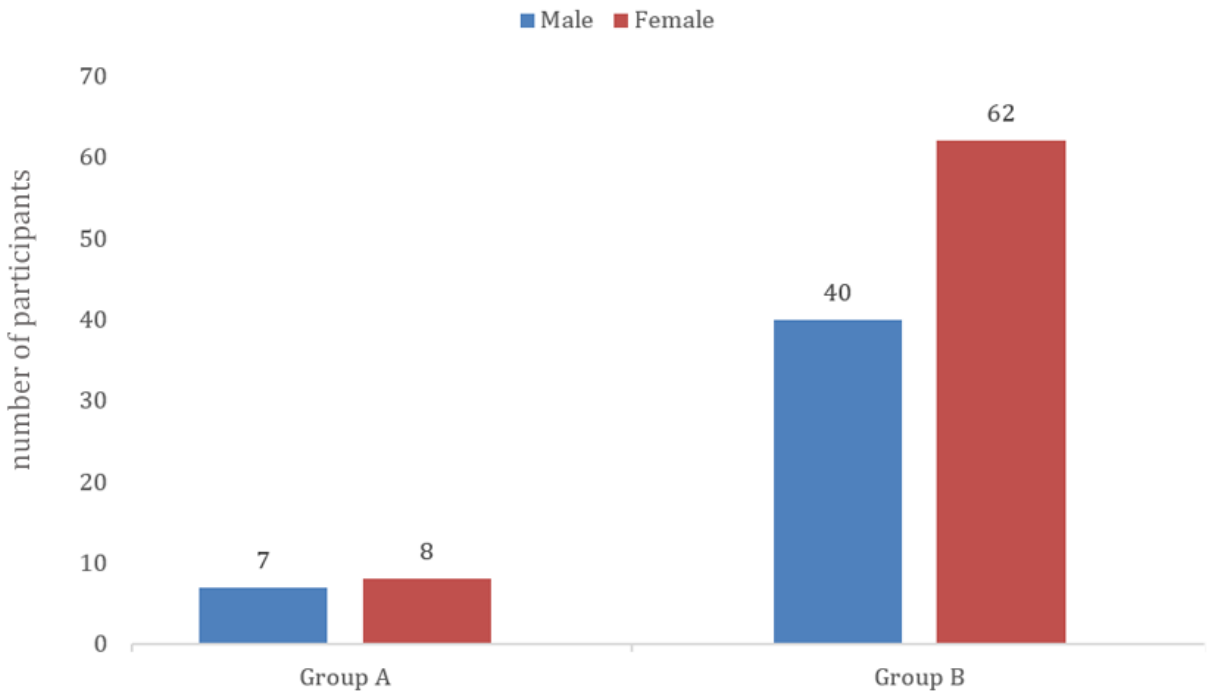
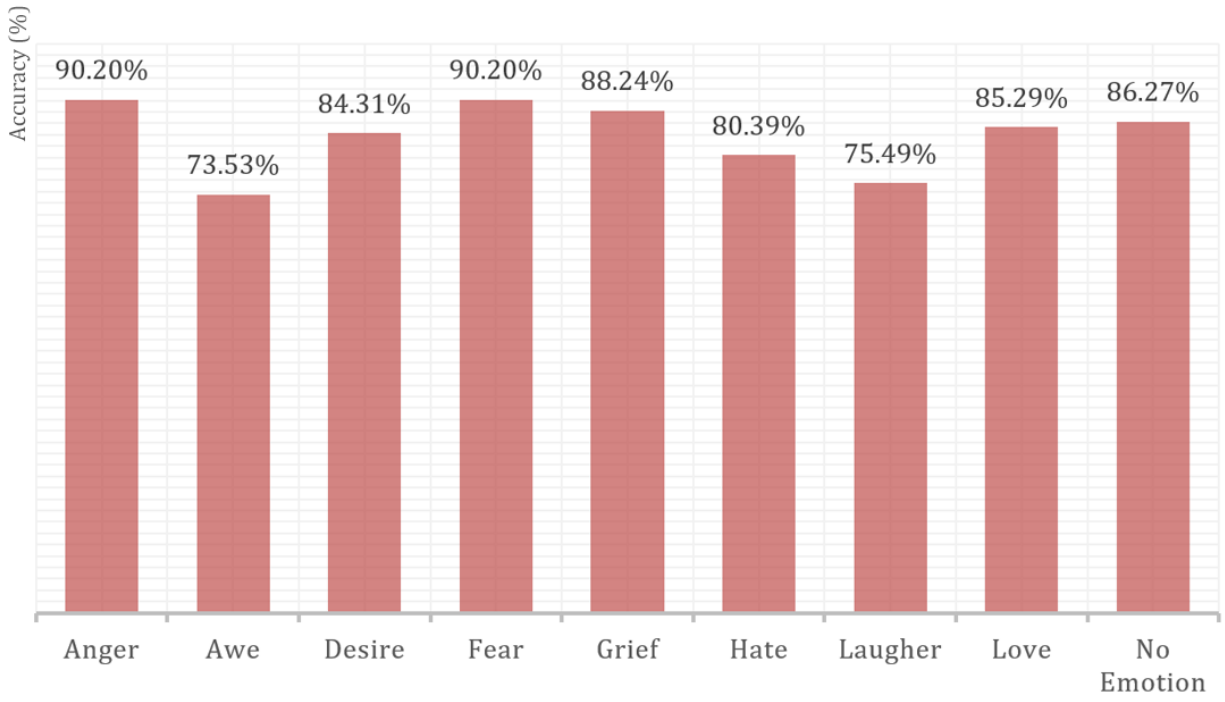


Table 3. Sensitivity tests on participants' smartphones.

Device model	Screen density (pixels/inch)	Force granularity	Area granularity	Sensitive enough?
Apple iPhone 6S	326	121	95	Yes
Huawei G610U20	220	120	32	Yes
Huawei Y635TL00	196	1	1	No
LGE-D802	424	30	12	Yes
Motorola XT1023	256	52	16	Yes
OnePlus A2003	401	1	5	No
Samsung G920A	576	1	41	No
Samsung I8530	233	15	18	Yes
Sony Xperia XZ	424	35	21	Yes
Xiaomi RNote3	403	48	14	Yes
Xiaomi Redmi4	294	1	1	No

Figure 5. Human recognition of emotions in force-expressive touches.



However, 49.0% (50/102) of the participants were 100% (450/450) accurate in recognizing all 8 emotions (and no emotion) and 25.5% (26/102) were 77.8% (182/234) accurate. Male and female participants did equally well. Figure 5 shows the results obtained for each emotion and Table 4 details the classification results for each emotion. This experiment confirmed that emotions modulate our fingers as we force-touch a sensitive surface, as well as that humans have a high ability to recognize emotions from the pattern of the emotion in the expressive force-touch.

We can see distinct patterns between emotions in terms of their beginning and end, as well as the shape, speed, acceleration, and space occupation of their respective expressions. Figure 7 shows more details in the variation of the amount of force and skin area for each emotion. The horizontal axis is the time in milliseconds; the vertical axis represents the variation of force from 0 to 1, and the dot size on the curves correlates with the variation of the skin area over time.

Data Visualization

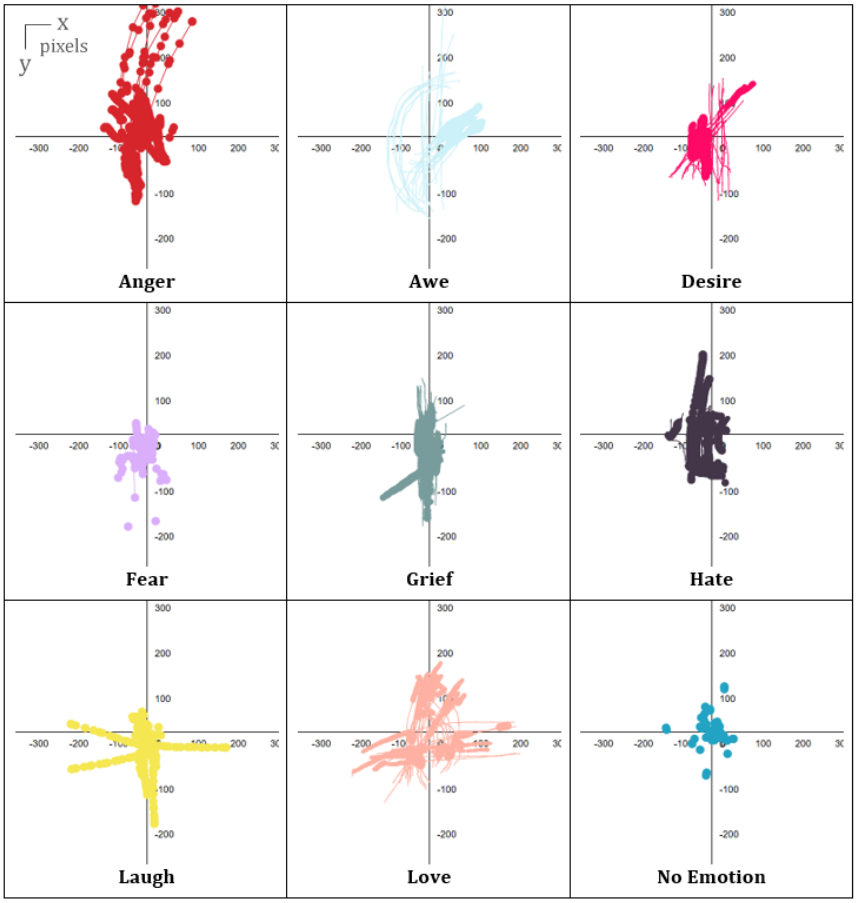
Figure 6 shows a single frame taken from participants' finger expressions of each emotion at a random instant t on the (x,y) axis.

Emotional expressions have distinct amounts of force and skin area. Hate is characterized by a very high amount of force and fear is very fast and short.

Table 4. Classification results of group B for each video.

Emotion expressed in the video	Participants' classification								
	Anger	Awe	Desire	Fear	Grief	Hate	Laughter	Love	No emotion
Anger	92	2	0	0	1	5	2	0	0
Awe	1	75	3	2	5	4	6	2	4
Desire	1	5	86	0	1	2	2	4	1
Fear	1	2	0	92	0	1	4	1	1
Grief	1	4	2	0	90	1	2	2	0
Hate	2	2	1	2	2	82	9	2	0
Laughter	4	2	2	2	2	3	77	4	6
Love	0	3	4	2	1	1	2	87	2
No emotion	0	7	3	2	0	1	1	0	88

Figure 6. Patterns of emotional expressions on the (x,y) axis.



Experiment 2: Machine Learning Classification

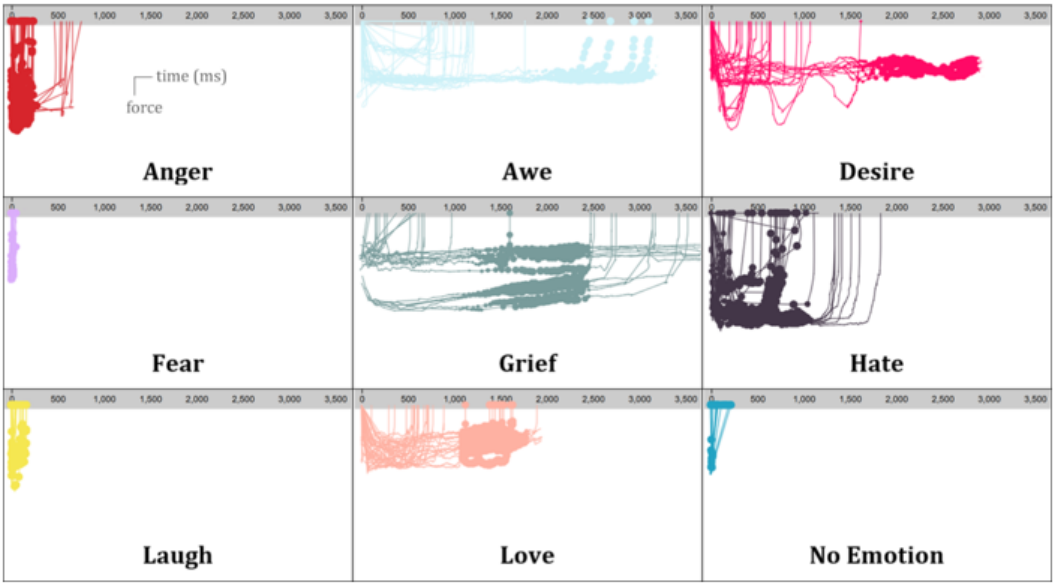
Table 5 shows the results of overall subjective ratings of group A for each emotion.

All participants expressed anger and no emotion very well (486/486, 100% of expressions were rated 5). Grief and laughter were the most challenging (20% of expressions: 25/126 for grief and 46/232 for laughter were rated below 3).

Our dataset comprised 2316 instances (or emotional expressions). Figure 8 shows the distribution of instances among emotions. Anger was the most expressed emotion with 486 expressions, and grief was the least expressed with 101 instances.

The number of touches per emotional expression varied between 1 and 25 (mean 1.94, SD 2.88); 83.46% (1933/2316) of expressions had an average of 1.38 touches and 3.45% (80/2316) of expressions had a maximum of 2.5 touches (see Figure 9). We saw more touches in laughter and anger than in the other emotions.

Figure 7. Force and skin variation of emotional expressions in time. The horizontal axis is the time in milliseconds; the vertical axis represents the variation of force from 0 to 1, and the dot size on the curves correlates with the variation of the skin area over time.



We designed an experimental framework in which we tested various machine learning techniques in supervised learning, including naive Bayes, nearest neighbor, neural networks, meta, and decision tree classifiers. We obtained the best classification results in random committee and 2 decision tree algorithms: random tree and random forest.

Random committee is a type of meta algorithm that takes classifiers and converts them into more powerful learners. Random committee builds an ensemble of base classifiers and averages their predictions [54]. Each one is based on the same data but uses a different random number seed. Decision trees are treelike structures; they start from root attributes and end with leaf nodes. Decision tree algorithms describe the relationships among attributes and the relative importance of attributes. Random tree chooses a test based on a given number of random features at each node, performing no pruning. Random forest constructs random forests by bagging ensembles of random trees [55]. Table 6 shows the classification results using the 10-fold cross-validation test option.

The percentage of correctly classified instances was very high, and varying between 86.14% (1995/2316) and 91.11% (2110/2316). Kappa statistics is a chance-corrected measure of agreement between the classifications and the true classes. It is calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. Kappa being higher than .81 demonstrates an almost perfect agreement for all the classifiers. Random forest produced the best results. Table 7 shows its detailed accuracy per emotion and Table 8 shows the confusion matrix.

The rate of true positives (or recall) varied from .75 for awe to 1.00 for fear. Most instances were correctly classified. The rate of false positive was insignificant and was highest in love, with only .03 instances falsely classified. The proportion of instances that were truly of a class divided by the total instances classified as that class (precision) varied from .82 to 1.00. A combined measure for precision and recall calculated as 2 × precision × recall / (precision + recall) is presented as the F measure. The area under the receiver operating characteristic curve approach 1.00 for all classes (>.98), which demonstrates the optimality of our model.

The confusion matrix in Table 8 shows the raw numbers, with anger, are, desire, fear, hate, grief, laughter, love, and no emotion being the class labels.

To test the performance of the classifier, we trained it on 15 subsets where we excluded 1 different participant in each run to be tested on the sample of the excluded participant (leave-one-run-out cross-validation and leave-one-sample-out cross-validation). The average performance of the classifier was 86.36%, kappa=.84: a drop of 4.6% compared with the result obtained with 10-fold cross-validation.

Human Versus Machines

In recognizing our emotions from finger force-touches, our algorithm did better than group B participants. Figure 10 shows a comparison of accuracy between our algorithm and group B participants in recognizing emotions.

For group B participants, anger was the easiest recognizable emotion, while awe and laughter were the most difficult to guess. Our algorithms were best in detecting fear, with almost 100% accuracy.

Table 5. Group A subjective ratings (range 0-5) for each emotion in experiment 2, by proportion giving that rating.

Emotion and rating	Participants who chose the rating, n (%)
Anger (n=486)	
5	486 (100)
Awe (n=181)	
4	23 (12.7)
5	158 (87.3)
Desire (n=302)	
2	39 (12.9)
3	60 (19.9)
4	82 (27.2)
5	121 (40.0)
Fear (n=292)	
4	79 (27.1)
5	213 (72.9)
Grief (n=126)	
0	25 (19.8)
4	25 (19.8)
5	76 (60.4)
Hate (n=303)	
4	61 (20.1)
5	242 (79.9)
Laughter (n=232)	
0	23 (9.9)
1	23 (9.9)
4	46 (19.8)
5	140 (60.4)
Love (n=288)	
4	95 (33.0)
5	193 (67.0)
No emotion (n=216)	
5	216 (100)

Figure 8. Overall number of expressions for each emotion.

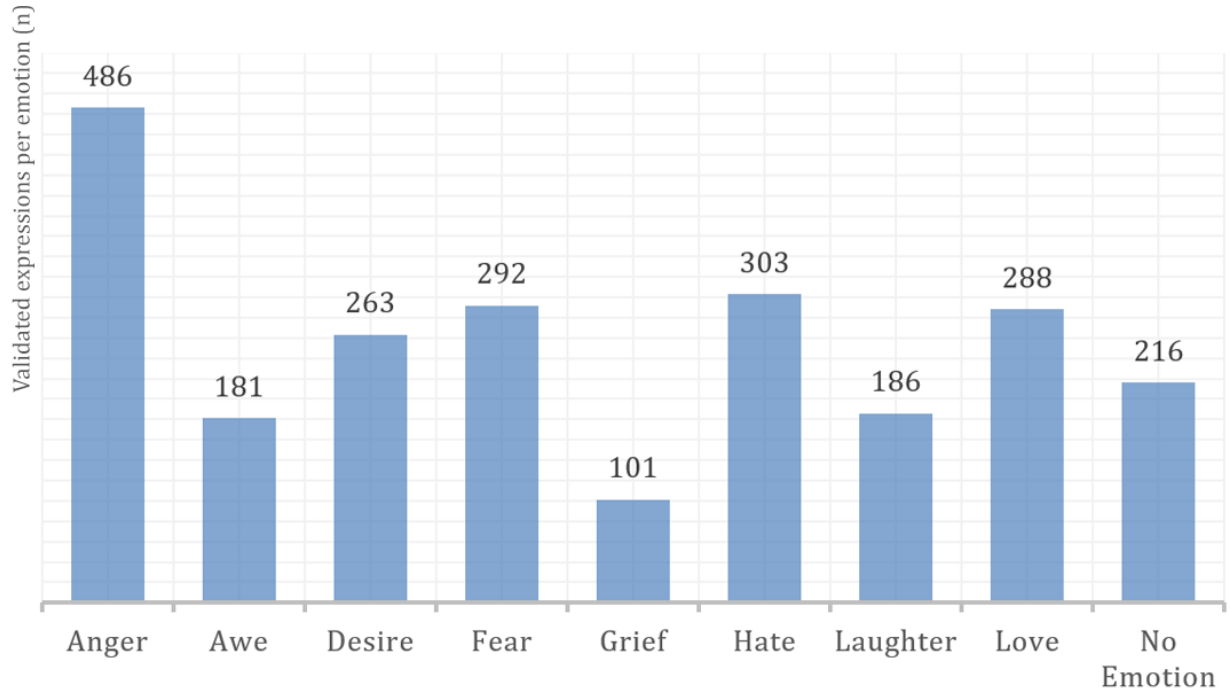


Figure 9. Number of touches per expression for each emotion.

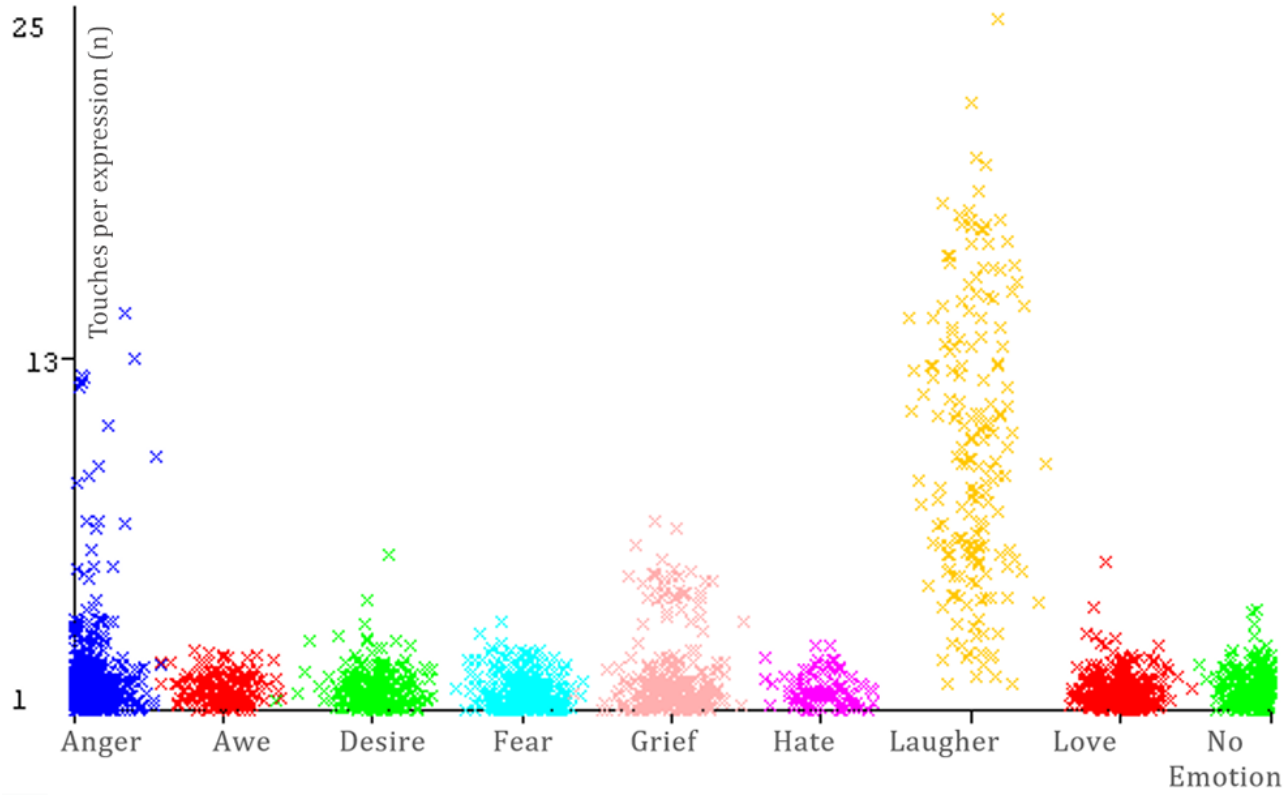


Table 6. Best emotion classification results using the 10-fold cross-validation test option.

Algorithm	Correctly classified, n (%)	Kappa statistic	Mean absolute error	Root mean square error	Relative absolute error (%)	Root relative square error (%)
Random tree	1995 (86.14)	.84	.03	.18	15.90	56.40
Random committee	2082 (89.90)	.88	.03	.13	16.43	42.62
Random forest	2110 (91.11)	.90	.04	.13	19.54	40.77

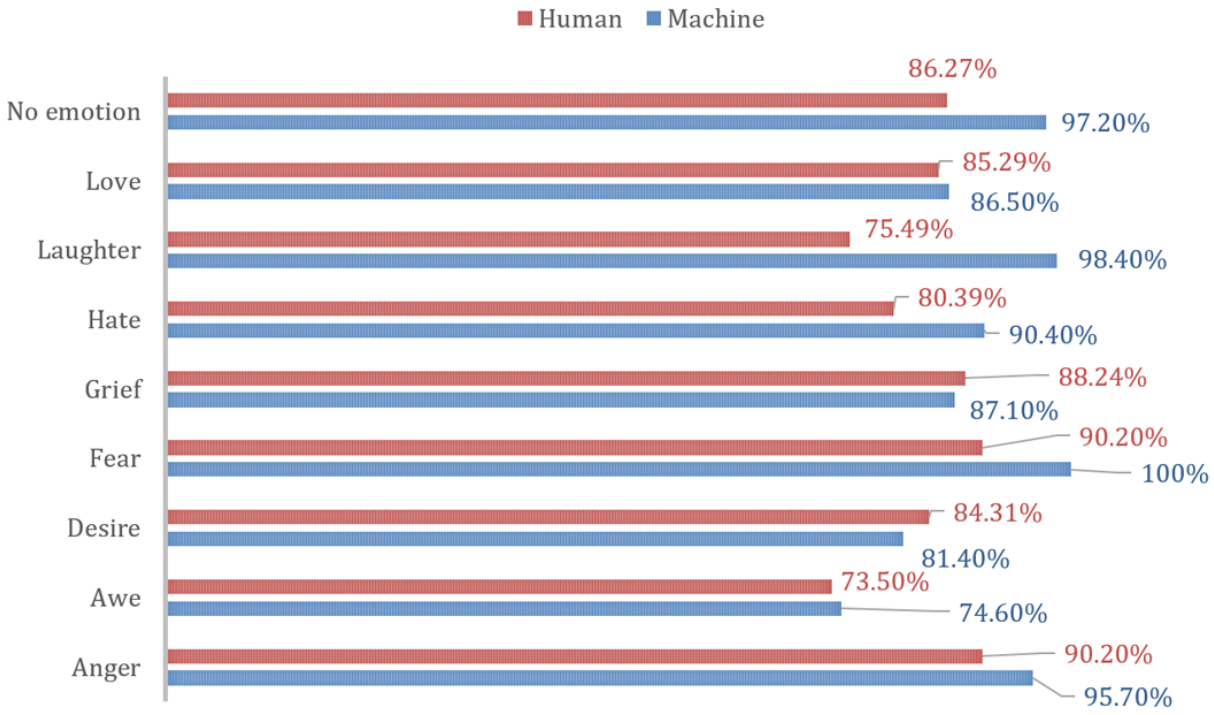
Table 7. Detailed accuracy per class for the random forest classifier.

Class	True positive rate	False positive rate	Precision	Recall	F measure	Matthews correlation coefficient	Area under the receiver operating characteristic curve	Area under the precision-recall curve
Anger	.957	.013	.951	.957	.954	.942	.994	.981
Awe	.746	.014	.818	.746	.780	.764	.980	.814
Desire	.814	.016	.866	.814	.839	.820	.980	.911
Fear	1.000	.000	1.000	1.000	1.000	1.000	1.000	1.000
Hate	.904	.016	.893	.904	.898	.883	.991	.940
Grief	.871	.008	.838	.871	.854	.848	.996	.926
Laughter	.984	.001	.989	.984	.987	.985	1.000	.999
Love	.865	.030	.803	.865	.833	.809	.978	.886
No emotion	.972	.003	.972	.972	.972	.969	.999	.995
Weighted average	.911	.012	.911	.911	.911	.899	.991	.946

Table 8. Confusion matrix for the random forest classifier.

Class	Anger	Awe	Desire	Fear	Hate	Grief	Laughter	Love	No emotion
Anger	465	5	1	0	0	0	2	10	3
Awe	0	135	12	0	0	5	0	29	0
Desire	2	4	214	0	28	1	0	13	1
Fear	0	0	0	292	0	0	0	0	0
Hate	3	2	14	0	274	7	0	3	0
Grief	3	3	0	0	2	88	0	5	0
Laughter	3	0	0	0	0	0	183	0	0
Love	10	15	5	0	3	4	0	249	2
No emotion	3	1	1	0	0	0	0	1	210

Figure 10. Human versus machine algorithm recognition of emotions.



Discussion

Principal Results

The results of the first experiment of this study demonstrated the ability of humans to recognize emotions when expressed through finger force-touch, and the results of the second experiment demonstrated clear finger force-touch patterning of emotions.

Our findings in experiment 1 indicated the ability of group B participants to recognize 8 emotions: anger, awe, desire, fear, grief, hate, laughter, love (and no emotion) as described in Table 1. Group B had a high accuracy of 83.8% (769/918) in recognizing the emotions of the emotionally expressive participants (group A) by only looking at the movement of their fingers when pressing a transparent glass to express emotions. Clynes [1] reported a similar ability of humans to recognize emotions by only looking at the movement in space of an arm expressing emotions. Clynes [1] and Hertenstein and colleagues [8,9] demonstrated the ability of humans to communicate and perceive distinct emotions via touch. Our finding in experiment 2 indicated higher accuracy for our algorithm, recognizing clear patterns of 8 emotions (and no emotion) with a 91.11% (2110/2316) classification accuracy. Following the training sessions, recording finger force-touch and skin area in the expression of each emotion and filtering the collected data based on subjective ratings revealed highly significant and accurate classification of group A's emotions.

An interesting insight in our data was the correlation between the 8 emotions (and no emotion) and the spatiotemporal features: coordinates of the beginning and end of an expression, the distance, angle, duration, velocity, and acceleration of the expression, the total number of touches, the mean duration of touches of an expression, and the mean, maximum, and velocity of both force and size of an expression.

Limitations

Using a strict participant selection process and a personalized emotion induction protocol with human validation and subjective rating allowed us to state that anger, awe, desire, fear, grief, hate, laughter, love, and no emotion produce specific response patterns in finger force-touch expression of emotions. However, there are several considerations (related to the selection process of group A participants) that limit generalization of our findings: group A participants (1) were highly emotionally aware, (2) had good imagery ability, (3) were used to smartphones, (4) had good touch dexterity on their smartphones, and (5) were willing and able to communicate emotions authentically using touch. Future replication of these findings is needed in participants who are poor imagers, less emotionally aware, and nonusers of smartphones.

Comparison With Prior Work

According to cognitive-physiological network models, ideas, memories, and verbal and subjective descriptors are very important components to induce emotions. Thus, when we scrupulously described our emotions to our participants and then asked them to remember and personalize meaningful emotion-evoking scenes and use their finger as the motor output to communicate the emotion to another participant before communicating the expression to a smartphone, the finger force-touch alone discriminated between anger, awe, desire, fear, grief, hate, laughter, love, and no emotion in 91.11% (2110/2316) of the cases. This indicates that the 9 spatiotemporal features extracted from finger force-touch measures were part of the network that was activated when participants expressed their emotions. Our accuracy rates were higher than those

reported in earlier similar studies where emotions or emotional dimensions were detected from force-touch, typing, and words or other sensors in the smartphone [28,32-35]. This may be related to clearer descriptions of our target emotions, the use of personalized multimedia material and imagery of real-life situations to induce emotions, and the use of another participant to validate the communication of the emotion via finger force-touch. Also, our higher accuracy may be due to the strict conditions in the selection process of participants for group A (they were all good imagers and highly emotionally aware).

Conclusions

Emotions are unique and important entities with built-in windows across the mind-body barrier that need to be understood. They convey great power in the development and mental healing of the individual, of society, and even for the now self-conscious evolution of human beings.

In this study, we described a protocol and implemented methods that allowed us to validate the human ability to express and perceive distinct emotions, and a machine's ability to recognize those emotions on force-sensitive smartphones. Much remains to be done to build more comprehensive, accurate, and scalable emotion detection algorithms in real-life contexts.

Touch is one of the most powerful of human senses, and we use it passively to communicate emotions. As we continue interacting with devices through touch, it is becoming essential to analyze these patterns and sense emotion. Enabling smartphone apps to capture, discern, and communicate the emotions of expressive touches will completely change the way users perceive and touch their devices and facilitate spontaneous emotional expressions. Human-computer-human interactions will get better, clearer, and emotionally intelligent.

Conflicts of Interest

AH is the founder, Chief Executive Officer, and Chief Technology Officer of Emaww Inc, a software startup based in Montreal, Canada. Emaww specializes in emotions, human-computer interactions, generative art, and artificial intelligence. It aims to build innovative solutions to empower users, professionals, and developers with emotional insight and improve human-computer-human communications.

References

1. Clynes M. Sentics: The Touch of Emotions. Garden City, NY: Anchor Press/Doubleday; 1977.
2. De Spinoza B. Ethics. London, UK: Penguin Classics; 1996.
3. Ardiel EL, Rankin CH. The importance of touch in development. Paediatr Child Health 2010 Mar;15(3):153-156 [FREE Full text] [Medline: 21358895]
4. Fairhurst MT, Löken L, Grossmann T. Physiological and behavioral responses reveal 9-month-old infants' sensitivity to pleasant touch. Psychol Sci 2014 May 01;25(5):1124-1131 [FREE Full text] [doi: 10.1177/0956797614527114] [Medline: 24681587]
5. Feldman R, Rosenthal Z, Eidelman AI. Maternal-preterm skin-to-skin contact enhances child physiologic organization and cognitive control across the first 10 years of life. Biol Psychiatry 2014 Jan 01;75(1):56-64. [doi: 10.1016/j.biopsych.2013.08.012] [Medline: 24094511]
6. Field T. Touch. Boston, MA: MIT Press; 2003.
7. Field T. Touch for socioemotional and physical well-being: a review. J Dev Rev 2010 Dec;30(4):367-383. [doi: 10.1016/j.dr.2011.01.001]
8. Hertenstein MJ, Keltner D. Gender and the communication of emotion via touch. Sex Roles 2011 Jan;64(1-2):70-80 [FREE Full text] [doi: 10.1007/s11199-010-9842-y] [Medline: 21297854]
9. Hertenstein MJ, Keltner D, App B, Bulleit BA, Jaskolka AR. Touch communicates distinct emotions. Emotion 2006 Aug;6(3):528-533. [doi: 10.1037/1528-3542.6.3.528] [Medline: 16938094]
10. Calvo R, Peters D. Positive Computing: Technology for Wellbeing and Human Potential. Cambridge, MA: MIT Press; 2014.
11. DeSteno D. Emotional Success: The Power of Gratitude, Compassion, and Pride. Boston, MA: Eamon Dolan/Houghton Mifflin Harcourt; 2018.
12. Feldman-Barrett L. How Emotions are Made: The Secret Life of the Brain. Boston, MA: Houghton Mifflin Harcourt; 2017.
13. Yonck R. The Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence. New York, NY: Arcade Publishing; 2017.
14. Doerrfeld B. 20+ Emotion recognition APIs that will leave you impressed, and concerned. Stockholm, Sweden: Nordic APIs AB; 2015. URL: https://nordicapis.com/20-emotion-recognition-apis-that-will-leave-you-impressed-and-concerned/ [accessed 2018-02-05] [WebCite Cache ID 6x0WJ4X4i]
15. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 2014 Dec;5(4):1093-1113. [doi: 10.1016/j.asej.2014.04.011]
16. Ekman P. The atlas of emotions. 2018. URL: http://atlasofemotions.org/ [accessed 2018-02-05] [WebCite Cache ID 6x0Z8E7eh]
17. Hjortsjo C. Man's Face and Mimic Language. Lund, Sweden: Studen Litteratur; 1969.

18. Neeru N, Kaur L. Performance comparison of recognition techniques to recognize a face under different emotions. Am J Eng Res 2016;5(12):213-217.
19. Doyle CM, Lindquist KA. Language and emotion: hypotheses on the constructed nature of emotion perception. In: Fernandez-Dols JM, Russell JA, editors. The Science of Facial Expression. New York, NY: Oxford University Press; 2017.
20. Doyle CM, Lindquist KA. When a word is worth a thousand pictures: language shapes perceptual memory for emotion. J Exp Psychol Gen 2018 Jan;147(1):62-73. [doi: 10.1037/xge0000361] [Medline: 29309197]
21. Chenchah F, Lachiri Z. Speech emotion recognition in noisy environment. 2016 Presented at: 2nd International Conference on Advanced Technologies for Signal and Image Processing; Mar 21-24, 2016; Monastir, Tunisia.
22. Heraz A, Frasson C. Towards a brain sensitive intelligent tutoring system: detecting emotions from brainwaves. Adv Artif Intell 2011;2011:1-13. [doi: 10.1155/2011/384169]
23. Kreibig SD. Autonomic nervous system activity in emotion: a review. Biol Psychol 2010 Jul;84(3):394-421. [doi: 10.1016/j.biopsycho.2010.03.010] [Medline: 20371374]
24. Levenson RW. The autonomic nervous system and emotion. J Emot Rev 2014;6(2):100-112. [doi: 10.1177/175407391351200]
25. Castellano G, Villalba SD, Camurri A. Recognising human emotions from body movement and gesture dynamics. 2007 Presented at: ACII 2007: Affective Computing and Intelligent Interaction; Sep 12-14, 2007; Lisbon, Portugal p. 71-82.
26. Kapur A, Babul NV, Tzanetakis G, Driessen PF. Gesture-based affective computing on motion capture data. In: Tao J, Tan T, Picard RW, editors. Affective Computing and Intelligent Interaction, First International Conference. Lecture Notes in Computer Science 3784. Cham, Switzerland: Springer; 2005:1-7.
27. Kipp M, Martin JC. Gesture and emotion: can basic gestural form features discriminate emotions? 2009 Presented at: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops; Sep 10-12, 2009; Amsterdam, Netherlands.
28. Dael N, Bianchi-Berthouze N, Kleinsmith A, Mohr C. Measuring body movement: currentfuture directions in proxemicskinesics. In: Matsumoto D, Hwang HC, Frank MG, editors. APA Handbook of Nonverbal Communication. Washington, DC: American Psychological Association; 2015.
29. Anil KK, Kiran B, Shreyas B, Sylvester J. A multimodal approach to detect user's emotion. Procedia Comput Sci 2015;70:2015-2303. [doi: 10.1016/j.procs.2015.10.096,]
30. Kapoor A, Picard RW. Multimodal affect recognition in learning environment. 2005 Presented at: 13th Annual ACM International Conference on Multimedia; Nov 6-11, 2005; Singapore p. 677-682. [doi: 10.1145/1101149.1101300]
31. Shah M, Chakrabarti C, Spanias A. A multi-modal approach to emotion recognition using undirected topic models. 2014 Presented at: IEEE International Symposium on Circuits and Systems (ISCAS); Jun 1-5, 2014; Melbourne, Australia p. 754-757.
32. Ghosh S. Emotion-aware computing using smartphone. 2017 Presented at: 9th International Conference on Communication Systems and Networks (COMSNETS); Jan 4-8, 2017; Bangalore, India.
33. Olsen AF, Torresen J. Smartphone accelerometer data used for detecting human emotions. 2016 Presented at: 3rd International Conference on Systems and Informatics (ICSAI); Nov 19-21, 2016; Shanghai, China.
34. Dai D, Liu Q, Meng H. Can your smartphone detect your emotion? 2016 Presented at: 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD); Aug 15, 2016; Changsha, China.
35. Sneha H, Annappa B, Manoj Kumar MV, Likewin T. Smart phone based emotion recognition and classification. 2017 Presented at: Second International Conference on Electrical, Computer and Communication Technologies; Feb 22-24, 2017; Coimbatore, India.
36. Number of mobile phone users worldwide from 2015 to 2020 (in billions). New York, NY: Statista, Inc; 2018.
37. Winnick M, Zolna R. Putting a finger on our phone obsession: mobile touches: a study on humans and their tech. Chicago, IL: dscout, Inc; 2016 Jun 16. URL: https://blog.dscout.com/mobile-touches [accessed 2018-02-05] [WebCite Cache ID 6x0ZVfTHG]
38. Nicholas M, Samuels M. The trust divide for brands online. London, UK: Kantar TNS; 2017. URL: http://www.tnsglobal.com/intelligence-applied/the-trust-divide-for-brands-online [accessed 2018-02-05] [WebCite Cache ID 6x0ZTcfZ8]
39. Grant J, Odlaug B, Chamberlain S. Why can't I Stop? Reclaiming Your Life From a Behavioral Addiction. Baltimore, MD: Johns Hopkins Press; 2016.
40. Higuchi S, Nakayama H, Mihara S, Maezono M, Kitayuguchi T, Hashimoto T. Inclusion of gaming disorder criteria in ICD-11: a clinical perspective in favor. J Behav Addict 2017 Sep 01;6(3):293-295 [FREE Full text] [doi: 10.1556/2006.6.2017.049] [Medline: 28816497]
41. Florence J. Mindfulness Meets Emotional Awareness, 7 Steps to Learn the Language of Your Emotions. 1st edition. London, UK: A-Z of Emotional Health Ltd; 2017.
42. Lang P, Bradley M, Cuthbert B. International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-6. Gainesville, FL: University of Florida; 2005.
43. Baveye Y, Bettinelli JN, Dellandréa E, Chen L, Chamaret C. A large video database for computational models of induced emotion. 2013 Presented at: ACII th International Conference on Affective Computing and Intelligent Interaction; Sep 2-5, 2013; Geneva, Switzerland.

44. Lane RD, Quinlan DM, Schwartz GE, Walker PA, Zeitlin SB. The Levels of Emotional Awareness Scale: a cognitive-developmental measure of emotion. J Pers Assess 1990;55(1-2):124-134. [doi: 10.1080/00223891.1990.9674052] [Medline: 2231235]
45. Betts G. The Distribution and Functions of Mental Imagery [doctoral thesis]. New York, NY: Teachers College, Columbia University; 1909.
46. Sheehan PW. A shortened form of Betts' questionnaire upon mental imagery. J Clin Psychol 1967 Jul;23(3):386-389. [Medline: 6082130]
47. Lane RD. Neural substrates of implicit and explicit emotional processes: a unifying framework for psychosomatic medicine. Psychosom Med 2008 Feb;70(2):214-231. [doi: 10.1097/PSY.0b013e3181647e44] [Medline: 18256335]
48. Barchard KA, Bajgar J, Leaf DE, Lane RD. Computer scoring of the Levels of Emotional Awareness Scale. Behav Res Methods 2010 May;42(2):586-595. [doi: 10.3758/BRM.42.2.586] [Medline: 20479190]
49. Lane RD, Sechrest L, Reidel R, Weldon V, Kaszniak A, Schwartz GE. Impaired verbal and nonverbal emotion recognition in alexithymia. Psychosom Med 1996;58(3):203-210. [Medline: 8771618]
50. Lane RD, Sechrest L, Riedel R, Shapiro DE, Kaszniak AW. Pervasive emotion recognition deficit common to alexithymia and the repressive coping style. Psychosom Med 2000;62(4):492-501. [Medline: 10949094]
51. Miller GA, Levin DN, Kozak MJ, Cook EW, McLean A, Lang PJ. Individual differences in imagery and the psychophysiology of emotion. Cognit Emot 1987 Oct;1(4):367-390. [doi: 10.1080/02699938708408058]
52. Coan J, Allen J. Handbook of Emotion Elicitation and Assessment. Series in Affective Science. Oxford, UK: Oxford University Press; 2007.
53. Peres-Neto PR, Jackson DA, Somers KM. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Comput Stat Data Anal 2005 Jun;49(4):974-997. [doi: 10.1016/j.csda.2004.06.015]
54. Witten I, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. 2nd edition. San Francisco, CA: Morgan Kaufmann; 2005.
55. Breiman L. Random forests. Mach Learn 2001;45:5-32.

Abbreviations

ANS: autonomic nervous system
LEAS: Levels of Emotional Awareness Scale
QMI: Questionnaire Upon Mental Imagery

Edited by R Calvo, M Czerwinski, G Wadley, J Torous; submitted 12.02.18; peer-reviewed by MH Yong, S Hussain; comments to author 03.04.18; revised version received 09.05.18; accepted 18.06.18; published 30.08.18.

Please cite as:
Heraz A, Clynes M
Recognition of Emotions Conveyed by Touch Through Force-Sensitive Screens: Observational Study of Humans and Machine Learning Techniques
JMIR Ment Health 2018;5(3):e10104
URL: http://mental.jmir.org/2018/3/e10104/
doi:10.2196/10104
PMID:30166276

Original Paper

Using Mobile Technology to Provide Personalized Reminiscence for People Living With Dementia and Their Carers: Appraisal of Outcomes From a Quasi-Experimental Study

Elizabeth A Laird[1], BSc (Hons), MSc, PhD; Assumpta Ryan[1], BSc (Hons), MEd, PhD; Claire McCauley[1], BPharm (Hons), MSc, PhD; Raymond B Bond[2], BCompSc (Hons), PhD; Maurice D Mulvenna[2], BSc (Hons), MPhil, PhD; Kevin J Curran[3], BSc (Hons), PhD; Brendan Bunting[4], BSc (Hons), MPsychol, PhD; Finola Ferry[4], BSc (Hons) Psychol, PhD; Aideen Gibson[2], BCompSc (Hons), MSc

[1]Institute of Nursing and Health Research, Ulster University, Derry, United Kingdom

[2]School of Computing, Ulster University, Belfast, United Kingdom

[3]School of Computing, Engineering & Intelligent Systems, Ulster University, Derry, United Kingdom

[4]Psychology Research Institute, Ulster University, Derry, United Kingdom

Corresponding Author:
Elizabeth A Laird, BSc (Hons), MSc, PhD
Institute of Nursing and Health Research
Ulster University
Northland Road
Derry, BT48 7JL
United Kingdom
Phone: 44 2871675006 ext 75006
Fax: 44 2871675493
Email: ea.laird@ulster.ac.uk

Abstract

Background: Dementia is an international research priority. Reminiscence is an intervention that prompts memories and has been widely used as a therapeutic approach for people living with dementia. We developed a novel iPad app to support home-based personalized reminiscence. It is crucial that technology-enabled reminiscence interventions are appraised.

Objective: We sought to measure the effect of technology-enabled reminiscence on mutuality (defined as the level of "closeness" between an adult living with dementia and their carer), quality of carer and patient relationship, and subjective well-being.

Methods: A 19-week personalized reminiscence intervention facilitated by a program of training and a bespoke iPad app was delivered to people living with dementia and their family carers at their own homes. Participants (N=60) were recruited in dyads from a cognitive rehabilitation team affiliated with a large UK health care organization. Each dyad comprised a person living with early to moderate dementia and his or her family carer. Outcome measurement data were collected at baseline, midpoint, and intervention closure.

Results: Participants living with dementia attained statistically significant increases in mutuality, quality of carer and patient relationship, and subjective well-being (P<.001 for all 3) from baseline to endpoint. Carers attained nonsignificant increases in mutuality and quality of carer and patient relationship and a nonsignificant decrease in subjective well-being.

Conclusions: Our results indicate that individual-specific reminiscence supported by an iPad app may be efficient in the context of early to moderate dementia. A robust randomized controlled trial of technology-enabled personalized reminiscence is warranted.

(JMIR Ment Health 2018;5(3):e57) doi:10.2196/mental.9684

KEYWORDS

dementia; evaluation; mobile apps; reminiscence; research; technology; mobile phone

Introduction

Background

Dementia is an umbrella term that encompasses at least 40 conditions that feature progressive cognitive decline and are more prevalent in older age. In tandem with international aging demographics, the prevalence of dementia and associated costs have risen substantially. The estimated annual UK cost of dementia is over £26 billion [1], and this is higher than the combined costs for cancer, stroke, and heart disease. There is increasing evidence that nonpharmacological interventions for the symptoms of dementia can have commensurate effectiveness to pharmacological treatment and may be preferable where medication can cause negative side effects [2-4]. The progressive nature of dementia presents a challenge for families providing care to a relative with this condition [5-7]. Not surprisingly, the World Health Organization (WHO) has prioritized dementia as a global public health concern and has recommended that more research should be undertaken to inform supportive interventions for people living with dementia and their families [8].

Reminiscence refers to a range of psychosocial interventions that prompt memories and has been widely used as a therapeutic approach for people living with dementia and their carers [9,10]. Technology-based reminiscence increases opportunities to participate in conversations and enhances the social interactions of people living with dementia and their carers [10]; furthermore, it enables remote reminiscence to be delivered at home [11]. Traditional reminiscence utilizes collections of resources such as memory boxes that can stimulate a range of senses, including touch, taste, and smell. In contrast, technology-based reminiscence is reliant on visual and auditory memory prompts. These limitations may be offset by the portability, mobility, and utility of technology-based reminiscence systems to deliver personalized reminiscence experiences.

A systematic review [12] of technology-supported reminiscence therapy identified 44 papers that met the selection criteria. Although limited by the small sample sizes of some of the studies, the authors concluded that there were benefits to using information and communication technology (ICT) for reminiscence interventions. These benefits include access to rich and engaging multimedia reminiscence materials [13,14], opportunities for people living with dementia to participate in social interactions and take ownership of conversations [15,16], and a reduction in motor deficit-related barriers when interacting with media [16,17]. In the abovementioned review [12], 10 reviewed papers reported on the use of “reminiscence kits” that featured a technological component. Audio was a major component of these reminiscence kits, but impact evaluation was not reported. One study [17] examined the attitudes of older people (N=19) toward using an iPad to aid reminiscence. Participants in the study were randomly allocated to reminisce using either an iPad or more traditional images and cards. The results from that study indicated that participants enjoyed using the iPad. In a follow-up mixed methods design, a mobile app called “Memory Matters” was developed to promote reminiscence [15]; 18 people living with dementia and 8 family carers were asked to use MM for a period of 4 weeks. Consistent with the findings of a more recent study that explored a similar device [18], the technology-supported reminiscence was favorably evaluated. Family carers enjoyed discussing the early years with their relative, and on several occasions, the people living with dementia shared memories in a direct response to prompts provided by MM. People living with dementia who had only interacted minimally, or who had never spoken before, were observed to interact and support each other while using the app. These findings support the social engagement potential of mobile devices in the context of family caregiving in dementia [19,20].

As dementia progresses, it is common for carers to report a “disappearance of the relationship” [21,22]. There is a need to support caregiving relationships in order to protect mental and physical well-being of the carers [23,24]. ICT has an important role to play in this endeavor by supporting social connectivity [25]. It is important, therefore, that in studies of technology-based reminiscence, family carers are included in addition to people living with dementia.

Our research team was motivated by the rising acceptability of health apps to develop and test the feasibility of a novel app to deliver personalized reminiscence among people living with dementia and their carers. Consistent with recommendations [12], validated and standard outcome measures were selected for the appraisal of efficacy. Mutuality is a scale that measures closeness in a relationship [26]; WHO-Five Well-Being Index (WHO-5) is a short scale for measuring emotional well-being [27], and Quality of Carer-Patient Relationship scale (QCPR) is a scale of family caregiving [28]. All 3 scales have been tested in dementia research, but not in reminiscence research. This paper contributes to the evidence base by reporting the preliminary efficacy of technology-based personalized reminiscence facilitated by a program of training and an iPad app on mutuality, quality of caregiving relationships, and subjective well-being among people living with dementia and their family carers.

Development of a Reminiscence App

The size, capacity, and low cost of ubiquitous mobile devices have made them an attractive option for technology-based reminiscence systems. As part of this study, a cross-platform device agnostic tablet app (called InspireD: an acronym for Individual-Specific Reminiscence for People living with Dementia) was developed to facilitate reminiscing activity. The two primary aims of the app were to enable people living with dementia and their family carers to select and store personalized memorabilia (photographs, videos, sounds, music, etc) and to provide easy access to these visual and audiovisual cues to support bespoke reminiscence.

The InspireD app was developed [29] with input from the Reminiscence Network Northern Ireland and a user development group that comprised a total of 7 dyads, with each dyad comprising a person living with dementia and his or her primary caregiver (n=14). The Agile software development approach [30] was adopted to allow the functional prototype to be created early in the development lifecycle, with testing for usability and refinement taking place throughout the development process

[29]. The app was implemented using Appcelerator Studio (Appcelerator Titanium SDK, US), an Eclipse-based integrated development environment that provides an environment to build, test, package, and publish apps for various platforms, including iOS and Android. The code is written in JavaScript, with native user interface (UI) elements being invoked at runtime. It incorporates local facilities for persistent data storage in SQLite database and facilitates the use of third-party app programming interfaces for Flickr and YouTube (Figure 1).

The app consists of a UI that is usable and responsive across a variety of mobile devices (tablets, mobile phones). It is also possible to use the system on a personal computer or laptop via the Web browser. The main user (and co-users, ie, carers) can upload images, videos clips, and audio clips to the app. SQLite database functionality is used to store and manage data natively.

The main UI consists of a simple screen for people living with dementia to upload files with help from a reminiscence trainer or a family caregiver. A multiscreen layout allows users to choose which memorabilia they wish to access: view photos, watch videos, listen to audio files, and browse selected resources (Figure 2).

The design is minimalist, using verbal descriptors as well as images and icons to reinforce and indicate functionality to the user. Data are organized and presented primarily in the form of on-screen menus. The welcome screen is a simple log-in screen where the users confirm their identity by clicking a photo of themselves. The user data are contained within a local SQLite database, which can be easily queried with the reporting services enabled. Multimedia reminiscing resources (photos, videos, and audios) are also stored locally in the app data directory.

Figure 1. InspireD app system architecture SQLiteDB SQLite database.

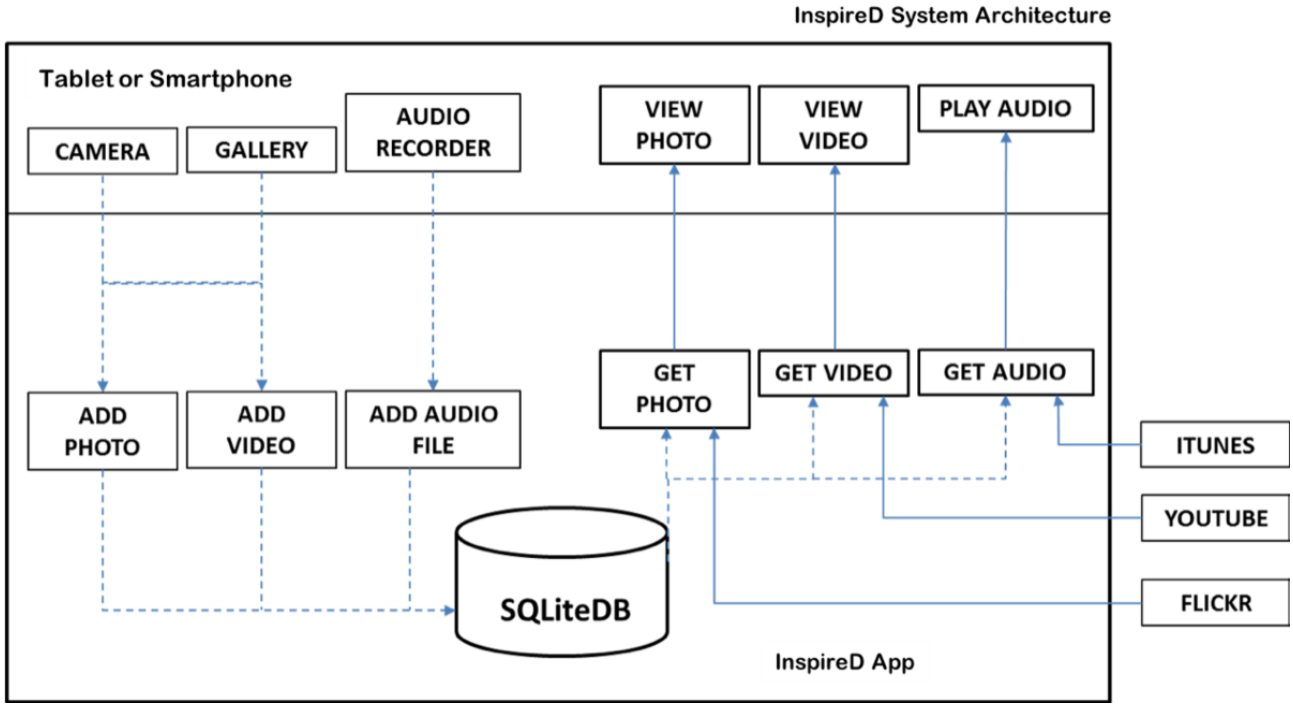
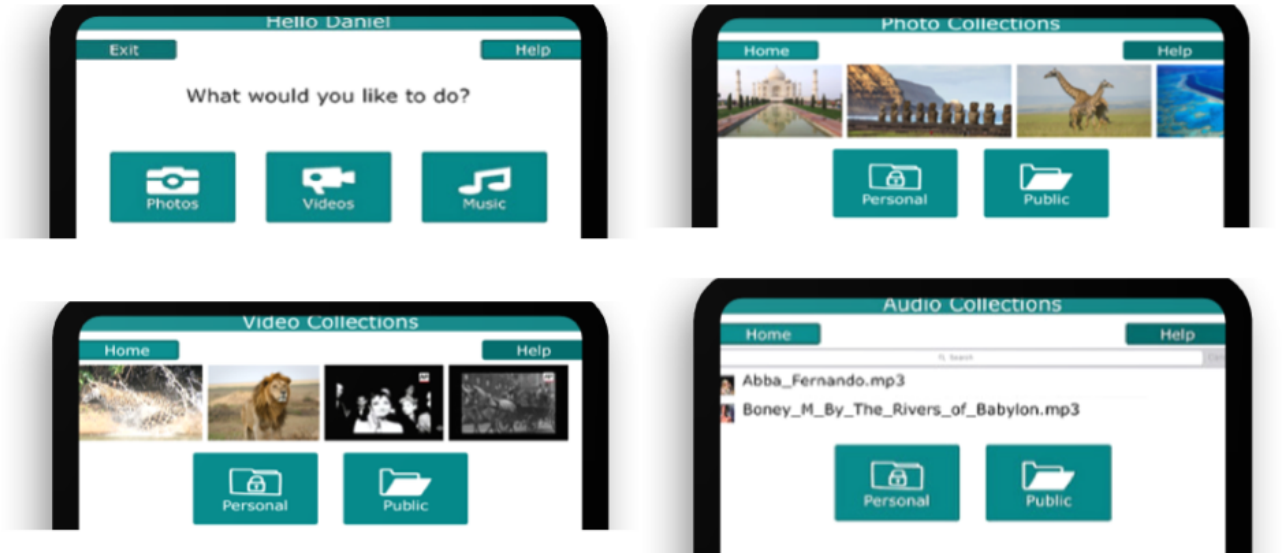


Figure 2. The InspireD reminiscence app user interface.



The InspireD app incorporates a logging facility for 5 canonical events. These are entry (logging in), admin (adding a photo, deleting an audio, etc), reminiscing (viewing a video, viewing a photo, etc), ecological momentary assessment (EMA) questions, and exit (logging out). Usage data across the course of the intervention were collected via secure email and statistically analyzed, and the findings have been published [31]. EMA is influenced by Kurt Lewin [32]. The use of "in-the-moment" approaches, along with rigorous measurement techniques in psychometric research, has been validated in recent research [33-36]. In our study, EMA involved the delivery of a small series of 5 questions directly to participants through the app; the feasibility of this approach in the context of dementia care is being appraised and will be reported in a subsequent paper. The InspireD system was designed with scalability in mind for future enhancements as it is envisaged that the final version will be a secure, cloud-based app accessible via a secure internet connection for authorized users. While the design incorporated the ability to store content locally on the device or to upload it to a cloud-based storage, a decision was made by the team that for the feasibility study, all content would be stored locally on the device.

Methods

Design

This paper reports on a feasibility study that incorporated a quasi-experimental design. An intervention of home-based personalized reminiscence, supported by a program of training and a novel iPad app, was examined for preliminary efficacy using 3 outcome measures pertaining to mutuality, emotional well-being, and quality of carer-patient relationships. The quasi-experimental design is appropriate and a common design for assessing the feasibility of novel technological interventions [37-39]. In line with quality standards, repeated measures testing was employed. Data were collected at baseline, midpoint, and intervention close. Table 1 outlines the intervention activities and data collection time-points.

The model of reminiscence that was utilized to underpin the training intervention was that of simple reminiscence [40], which encompasses mainly unstructured autobiographic storytelling and triggers that generate spontaneous reminiscence, often within a relational context, such as special days or events shared by friends and family. The goal of this approach is to enhance social contacts and short-term well-being while also supporting intergenerational bonding [40,41]. Our intervention was designed to cater for the needs, preferences, and interests of people living with dementia and their family carers. As a measure of quality, the reporting of this study adheres to the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement [42].

Settings

The setting was a large health and social care trust in a region of the United Kingdom. The trust catchment area is a mix of rural and urban communities serving a population of approximately 300,000 people, with an estimated 2717 of them living with dementia. Recruitment was facilitated by the trust's community mental health team for older people and through the trust's cognitive rehabilitation team as engagement with the latter was indicative of a diagnosis of early to moderate dementia.

Participants

A purposive sampling strategy was used to recruit 30 caregiving dyads (30 persons living with dementia and 30 carers). A sample size of 40-50 is recommended as sufficient for a feasibility study to estimate the total sample size across parameters to inform a future randomized controlled trial (RCT) [43]. Our participants were predominantly older people, and there was potential for significant dropout. With that in mind, we increased our sample size to 60.

Inclusion and exclusion criteria were developed to minimize the potential for bias in the recruitment process. We included people who (1) had a diagnosis of early to moderate dementia, (2) were able to communicate and understand conversations, and (3) were aware of their dementia diagnosis; furthermore, we included family carers who were (1) aged ≥18 years, (2) caring for a family member living with dementia meeting the above criteria (either cohabiting or non-cohabiting), and (3) aware of their relative's dementia diagnosis. Individuals with a major illness or disability that hindered their ability to engage in the study were excluded. Recruitment was commenced in April 2016 and continued until the sample size of 30 dyads (N=60) was achieved in October 2016.

Ethical considerations principally pertained to voluntariness, supporting separate informed consent for the people living with dementia and their carers, handling and storage of data, and right to withdraw from the study. The study received ethical approval (REC Ref 16/NI/0002) in line with regional and National Health Service Trust research governance.

Primary Outcome Measure

The primary outcome measure was mutuality, defined as the positive quality of the relationship between the carer and the care recipient [26]. The Mutuality scale consists of 15 items. A sample item includes "How attached are you to him or her?" A 5-point scale is used, ranging from 0 (not at all) to 4 (a great deal). Higher scores indicate a higher level of mutuality, which may support relationships in difficult circumstances. The Mutuality scale has been tested for validity in previous studies of family caregiving and has demonstrated internal consistency [44,45]. The mean was calculated across the response scores for data analysis [45,46].

Secondary Outcome Measures

The secondary outcome measures comprised QCPR [28] and WHO-5 [27,47]. The QCPR is a 14-item scale measuring relationship quality, including level of warmth and level of criticism. The scale has demonstrated good internal consistency and concurrent validity with other measures of relationship quality and carer stress [48]. Responses are rated using a 5-point Likert scale, scored from 1 (totally disagree) to 5 (totally agree). The 6 items measuring criticism and conflict are reverse scored in computation [28]. Total QCPR scores are utilized for data analysis. A total score >42 is indicative of a good relationship.

The WHO-5 comprises 5 questions that tap into the subjective well-being of participants; it has been extensively tested for validity [49,50] and reliability [51,52] scale items are scored from 0-5 and then totaled, giving a potential raw score ranging from 0 to 25. It is recommended that in studies assessing change over time, the WHO-5 raw scores are transformed to percentage scores for data analysis [53]. A total percentage score of ≤50 is an indication of low mood, and a score of ≤28 suggests likely depression, warranting further assessment.

Intervention and Follow-Up

Participating dyads were provided with a new touch screen tablet device that hosted the novel InspireD reminiscence app. A package of 5 reminiscence training sessions influenced by evidence-informed guidelines [40,41] was delivered by a reminiscence trainer employed by the Reminiscence Network Northern Ireland. Three information technology (IT) training sessions were then provided by an IT assistant to support the participants in uploading their personal memorabilia and to use the reminiscence app independently. The reminiscence and IT training packages were provided face-to-face, at the homes of participants living with dementia. The estimated cost of the intervention, which included the training package together with the cost of the InspireD system, was £2570 per dyad. After training was completed, participants were requested to engage in simple reminiscing through the app 3 days per week for the following 12 weeks. Compliance was supported by a user-friendly instruction booklet. A phone number for the IT trainer was provided, should any technological issues arise.

Participants were followed up for a period of 19 weeks from baseline (T_0), as outlined in Table 1. Midpoint measurement data (T_1) were collected on week 13 from baseline, which was 6 weeks into the independent use of the reminiscence technology. Endpoint measurement data were collected on week 19 at closure of the intervention (T_2). The data collection period was May 2016 to February 2017. Given the study design, there was no control group, and given the nature of the intervention, it was not possible to blind the participants or the trainers administering the intervention. All data were uploaded to IBM SPSS version 23 [54] using unique anonymized identification codes by the research assistant. The researcher responsible for analyzing the data and interpreting the results used that anonymized dataset.

Statistical Analysis

We used descriptive statistics to describe and synthesize the data pertaining to the characteristics of the participants. Missing data analysis, as recommended, was undertaken to discern possible patterns and challenges in the selected measurement tools [55]. We performed chi-square tests for analysis of nominal variables and independent *t* tests to compare measurement scores in Mutuality, QCPR, and WHO-5 across the dyad relationship (people living with dementia and their carers) and across gender at baseline; chi-square tests were not performed when numbers in categories were less than 5. Paired *t* tests were performed to investigate the differences in scores across 2 time-points. Furthermore, we performed within and between repeated measures analysis of variance (ANOVA) to investigate the impact of the intervention over time.

Table 1. Intervention activities and data collection time-points.

Timescale and activities	Data collection time-points	Repeated measures
Preintervention		
Baseline	Baseline (T_0)	• Mutuality • WHO-5[b] • QCPR[c]
19-week intervention		
Weeks 1-6 training		
Reminiscence training package (5 sessions)	N/A[a]	• N/A
Information technology training package (3 sessions)	N/A	• N/A
Week 7-19 reminiscence using iPad app		
Home reminiscence begins in week 7	N/A	• N/A
Home reminiscence continues in week 13	Midpoint (T_1)	• Mutuality • WHO-5 • QCPR
Home reminiscence ends in week 19	Endpoint (T_2)	• Mutuality • WHO-5 • QCPR

[a]N/A: not applicable.

[b]WHO-5: World Health Organization–Five Well-Being Index.

[c]QCPR: Quality of the Carer and Patient Relationship scale.

Correlational tests were performed to investigate relationships between continuous variables. On an intention-to-treat basis [55,56], missing data for Mutuality, QCPR, and WHO-5 were managed using the expectation-maximization imputation approach.

Results

Baseline Assessment

A baseline assessment of demographic details, Mutuality, QCPR, and WHO-5 was conducted prior to the program of reminiscence and IT training. We recruited 60 participants, in 30 dyads, in this study. Of them, a total of 58 participants (29 dyads) were retained in the study at completion. The participant characteristics and baseline measurement scores are presented in Table 2. Of all the participants with dementia, 67% (20/30) were men. A chi-square test for independence (with Yates continuity correction) was performed, and it revealed that the gender composition of the carer participants was different from that of participants living with dementia. Of all the carers, 80% (24/30) were women (P=.001). The age range of participants living with dementia was 61-94 years and that of the carers was 31-91 years. An independent t test revealed that the age of the carers (mean 67 years, SD 14.8) was significantly lower than that of the participants living with dementia (mean 79 years, SD 8.9; P<.001). In 23 of the dyads, the carer was living in the same house as the participant living with dementia. The majority of the participants living with dementia had little or no IT experience, whereas the majority of carer participants had at least moderate IT experience.

There were no missing data at baseline. Mean mutuality score at baseline was 3.13 (SD 0.68), indicating a moderate level of closeness in the relationship. Visual inspection of the histogram and Q-Q plots indicated a reasonable but positively skewed distribution.

Table 2. Baseline characteristics of participants.

Characteristic	Number of participants (N=60)	People living with dementia (n=30)	Family carers (n=30)	P value[a]
Age (years), mean (SD)	73 (13)	79 (8.9)	67 (14.8)	<.001
Age range (years)	31-94	61-94	31-91	N/A[b]
Gender, n (%)				
Male	26 (43)	20 (67)	6 (20)	N/A
Female	34 (57)	10 (33)	24 (80)	.001
Marital status, n (%)				
Married	47 (78)	22 (73)	25 (83)	N/A
Widowed	9 (15)	8 (27)	1 (3)	N/A
Separated or single	4 (6.7)	0 (0)	4 (13)	N/A
IT[c] experience, n (%)				
Little or none	35 (58)	24 (80)	11 (37)	N/A
Moderate	21 (35)	5 (17)	16 (53)	N/A
A lot	4 (7)	1 (3)	3 (10)	N/A
Home internet access, n (%)	52 (87)	25 (83)	27 (90)	N/A
Hobby choices, n (%)				
Social	29 (48)	14 (47)	15 (50)	N/A
Physical fitness	19 (32)	8 (27)	11 (37)	N/A
Creative	7 (12)	4 (13)	3 (10)	N/A
No hobby	5 (8)	4 (13)	1 (3)	N/A
Repeated measures, mean (SD)				
Mutuality	3.13 (0.68)	3.24 (0.54)	3.02 (0.79)	.22
QCPR[d]	57.4 (7.9)	58.1 (7.1)	56.7 (8.6)	.52
WHO-5[e]	61.0 (23.9)	60.8 (26.2)	61.2 (21.8)	.94

[a]P values <.05 indicate significance.

[b]N/A: not applicable.

[c]IT: information technology.

[d]QCPR: Quality of Carer-Patient Relationship scale.

[e]WHO-5: World Health Organization–Five Well-Being Index.

There was no significant difference in baseline mutuality for participants living with dementia and carers (*P*=.22). There was a statistically significant difference between mutuality for men (mean 2.9, SD 0.78) and women (mean 3.3, SD 0.56), with women having higher scores (95% CI −0.756 to −0.026; *P*=.04). No relationship was discerned between age and baseline mutuality (r=0.09).

Mean QCPR score was 57.4 (SD 0.9), indicating a good relationship. A Kolmogorov-Smirnov statistic of 0.057 was suggestive of a normal distribution, supported by visual inspection of histograms and Q-Q plots. There was no significant difference in baseline QCPR score between participants living with dementia and carers (*P*=.52) or between men and women (*P*=.91). There was a positive but weak correlation between age and QCPR (r=−0.123), which failed to reach statistical significance (*P*=.35).

Mean WHO-5 score was 61.0 (SD 23.9), indicating a moderate level of subjective well-being. The Shapiro-Wilks statistic (appropriate for a small sample) of 0.052 suggested normality, and this was supported by visual inspection of histograms and Q-Q plots. There was no significant difference in WHO-5 score between participants living with dementia and carers (*P*=.94) or between men and women (*P*=.40). No relationship was discerned between age and baseline WHO-5 (r=−0.04).

Missing Data Analysis

At T_1 (midpoint), 5% (3/60) participants had missing data: participant 50 had 2.9% missing data; participant 11 had 5.9% missing data, and participant 44 was unavailable for data collection due to a hospital admission. At T_2 (endpoint), 6.6% (4/60) participants had missing data: participant 20 had 2.9% missing data; participant 43 was unavailable due to a hospital admission, and one participant had died and her carer withdrew from the study.

Intention-to-Treat Analysis

Paired-samples *t* tests were conducted to compare baseline and endpoint measurement scores. There was a statistically significant increase in mutuality scores of participants living with dementia from baseline (mean 3.24, SD 0.545) to endpoint (mean 3.64, SD 0.274; 95% CI −0.56 to −0.23; *P*<.001, two-tailed). Furthermore, there was a statistically significant increase in QCPR scores of participants living with dementia from baseline (mean 58.07, SD 7.12) to endpoint (mean 63.2, SD 4.32; 95% CI −7.42 to −2.84; *P*<.001, two-tailed). Similarly, there was a statistically significant increase in WHO-5 scores of participants living with dementia from baseline (mean 60.8, SD 26.2) to endpoint (mean 70.6, SD 21.4; 95% CI −14.8 to −4.84; *P*<.001, two-tailed).

Regarding the carers, there was an increase in mutuality from baseline (mean 3.02, SD 0.79) to endpoint (mean 3.07, SD 0.60), but the increase was not statistically significant (*P*=.52). Similarly, an increase in carer QCPR scores from baseline (mean 56.7, SD 8.66) to endpoint (mean 57.9, SD 8.26) was not statistically significant (*P*=.28). There was a decrease in carer WHO-5 scores from baseline (mean 61.2, SD 21.8) to endpoint (mean 60.2, SD 23.4), but the change in the scores was not statistically significant (*P*=.74).

Mixed between-within subjects ANOVA was performed to assess the impact of the reminiscence intervention over time and between participants living with dementia and the carers. For the participants living with dementia, mean mutuality increased from baseline to midpoint and then further increased at endpoint. For the carers, mean mutuality peaked at midpoint. A similar pattern for participants living with dementia and carers was observed in mean QCPR scores over time. Regarding WHO-5, the mean scores of participants living with dementia increased from baseline to midpoint, with a further increase at endpoint. For the carers, mean WHO-5 scores decreased from baseline to midpoint and then increased to the end point. Mean outcome measurement scores, SDs, time-related *P* values, and pattern difference *P* values are presented in Table 3.

A statistically significant effect of the intervention on mutuality was demonstrated over time, Wilks lambda=0.77, F(2,57)=8.17, *P*=.001, partial eta squared=0.22. The pattern of mutuality scores for the participants living with dementia and the carers was significantly different, Wilks lambda=0.87, F(2,57)=4.23, *P*=.02, partial eta squared=0.129, with participants living with dementia exhibiting higher scores. A statistically significant effect of the intervention on QCPR was demonstrated over time, Wilks lambda=0.777, F(2,57)=8.15, *P*=.001, partial eta squared=.223. The pattern of QCPR scores for participants living with dementia and carers was statistically significant, Wilks lambda=0.88, F(2,57)=3.72, *P*=.03, partial eta squared=0.116. The participants living with dementia exhibited higher scores. Overall, the intervention did not demonstrate a significant effect on WHO-5 scores over time, Wilks lambda=0.90, F(2,57)=2.94, *P*=.06, partial eta squared=0.09. However, a statistically significant difference was found in the pattern of scores between the participants living with dementia and the carers, Wilks lambda=0.85, F(2,57)=4.90, *P*=.01, partial eta squared= 0.147. The participants living with dementia exhibited a pattern of higher scores.

Estimation of Sample Size for a Future Randomized Controlled Trial

A linear mixed model for a 2-way repeated measures ANOVA (fixed effects) was used to analyze the data. The between effect was dyad role, that is, participants living with dementia versus carers. The statistical power for the between effect in the model, based on the results from the mutuality measure, was 36 individuals per group (total=72). The power to detect the effects was set at 0.9 in all of the analyses.

For the within effect (repeated measures for both carers and those living with dementia), to detect the main effect of time (within subject effect), a sample of 16 respondents would be required in each group (total=32). For between-within subjects (interaction), to detect the interaction of condition (carer vs those living with dementia) and time, a sample of 39 individuals in each condition (total=78) would be required to detect an effect similar to that present in the previous study, with a statistical power of 0.9.

Table 3. Measures across time.

Measure and participant	Baseline (T_0), mean (SD)	Midpoint (T_1), mean (SD)	Endpoint, (T_2), mean (SD)	N	Time *P* value	Pattern *P* value
Mutuality	3.1 (0.687)	3.4 (0.55)	3.4 (0.543)	60	.001	N/A
PLWD[a]	3.2 (0.556)	3.6 (0.203)	3.6 (0.275)	30	N/A[b]	N/A
Carer	3.0 (0.799)	3.1 (0.67)	3.1 (0.601)	30	N/A	.02
QCPR[c]						
PLWD	58.1 (7.1)	61.3 (5.2)	63.2 (4.3)	30	N/A	N/A
Carer	56.8 (8.7)	58.6 (7.4)	58.9 (8.3)	30	N/A	.03
Total	57.4 (7.9)	59.9 (6.5)	60.6 (7.0)	60	.001	N/A
WHO-5[d]						
PLWD	60.8 (26)	69.9 (18)	70.7 (21)	30	N/A	N/A
Carer	61.2 (22)	56.5 (27)	60.3 (23)	30	N/A	.01
Total	61.0 (24)	63.2 (24)	65.5 (23)	60	.06	N/A

[a]PLWD: person living with dementia.

[b]N/A: not applicable.

[c]QCPR: Quality of Carer-Patient Relationship scale.

[d]WHO-5: World Health Organization–Five Well-Being Index.

Discussion

Principal Findings

In this quasi-experimental study, we sought to appraise outcomes from a feasibility study of individual-specific reminiscence facilitated by a program of training and an iPad app. A total of 58 participants (29 dyads) were retained in the study at completion, supporting the understanding that neither age nor a diagnosis of dementia are barriers to engagement in home-based research using technology. The main findings from the study are as follows: (1) statistically significant increases in mutuality, quality of carer and patient relationship, and emotional well-being of participants living with dementia from baseline to endpoint; (2) nonsignificant increases in mutuality and QCPR and a nonsignificant decrease in WHO-5 scores for the carer participants from baseline to endpoint; and (3) statistically significant differences in patterns of intervention effect across time, with the participants living with dementia exhibiting patterns of higher scores. It is difficult to determine the clinical significance of these changes as this was outside the remit of this feasibility study. However, a future RCT could include additional scales such as the mini-mental state examination [57] and Geriatric Depression Scale Short Form [58] in repeated measures testing.

Comparison With Prior Work

There is increasing use of mobile computer software and rising acceptability of health promotion apps internationally [29,59-62]. It was foreseen that the future would bring opportunities for reminiscing facilitated by touch screen interfaces [10]. Ours was a feasibility study with the objectives of testing a novel intervention of individual-specific reminiscence and investigating its impact. The total number of participants (N=60) recruited to this study is a significant increase from previous technology-enabled reminiscence studies [12]. The 3 outcome measures utilized in our study were mutuality [26], Quality of the Carer and Patient Relationship [28], and WHO-5 [47]. A strength of our study is that all 3 tools have previously undergone extensive testing for validity and reliability and were sufficiently sensitive to deliver statistically significant results. Our reminiscence intervention differed in a number of ways from the approaches taken in recent studies [12]: (1) our intervention was home based; (2) the participating dyads received a program of individual-specific training in reminiscing and IT, and (3) the reminiscing activity was supported by an iPad app hosted on tablet software with each of the participants having his or her own unique access log-in details. Our findings suggest that technology-based reminiscence may be able to support mutuality and quality of informal caregiving relationships, in contrast with negative trends observed in longitudinal studies among caregiving dyads [22,44]. We cannot make direct comparisons between our results and those of other technology-based reminiscence research due to the lack of appraisal of outcome in previous research. Our findings, however, add to the emerging evidence that technology-based reminiscence offers benefits in the context of family caregiving in dementia [12,15,18].

It is acknowledged internationally that family carers are the most important practical, personal, and economic supports for people with dementia [7,8] and that enduring caregiving roles in the context of dementia are associated with significant negative trends in mutuality and quality of life among family carers [22]. It would, therefore, not have been surprising if we had found significant negative trends in the outcome scores of carers over the course of our study. Our decision to deliver a home-based intervention was informed by research that suggested that research participation can pose a significant challenge to carers of people living with dementia [48]. Our

research findings suggest that it is possible that the home-based nature of our intervention contributed to the statistically significant enhancements for participants with dementia, with no significant detriment to carers. To what extent carers would continue to support the intervention in a longer-duration study is worthy of consideration. It may be possible to develop the iPad app further to utilize a coaching companion to prompt, incentivize, and reward the carer. Taken together, our findings support the need for a robust RCT of home-based app-enabled personalized reminiscence. A stratified sampling strategy guided by mini-mental state examination scores, with matched controls, and a longer follow-up time of up to 2 years would address the unknown issue of how long the intervention effect might last.

Limitations

This was a feasibility study, and it was important to maximize the exposure to the novel technology-based personalized reminiscence intervention. There are acknowledged challenges in the recruitment of people living with dementia and their family carers, and use of a comparison group would have reduced exposure to the intervention. Quasi-experimental designs, such as the one we adopted, cannot establish cause-and-effect relationships with certainty, but they can establish strong links. We cannot rule out a Hawthorne effect given the trial design and the possibility that pre-existing factors have influenced the results. Conclusions, therefore, have to be interpreted with caution. An additional limitation is the underrepresentation of women among the participants living with dementia in our study, given that women have been constituted as a marginalized majority in UK prevalence of dementia [63].

Conclusions

Reminiscence has been promoted internationally as a means of enhancing standards of care and quality of life of people living with dementia and their family carers. Our study comprised a novel intervention of home-based reminiscence with repeated measures testing. The findings of this study indicate statistically significant enhancements in mutuality, quality of relationship, and subjective well-being for the participants living with dementia and nonsignificant enhancements in mutuality and quality of relationship for carers. These findings support an emerging body of evidence that purports that individual-specific psychosocial interventions have efficacy in the context of dementia. It is important to highlight that our study is not without limitations and that pre-existing factors may have influenced the results. Nonetheless, our intervention, comprising a program of training and use of a novel iPad app, may contribute to the ongoing development of home-based reminiscence in the context of dementia. Future research must be cognizant of the potential for women living with dementia to be underrepresented among participants and the importance of controlling pre-existing factors. A robust RCT of personalized reminiscence is worthy of consideration.

Acknowledgments

This work was funded by the Public Health Agency Research and Development Office and Atlantic Philanthropies (COM/5016/14). The authors would like to thank the Reminiscence Network Northern Ireland and people living with dementia and their carers for their input in the cocreation and design of the InspireD app and the Trust's Cognitive Rehabilitation Team for facilitating recruitment. Special thanks to Alex Turnbull from Kainos for invaluable assistance in early-stage technology selection decisions.

Authors' Contributions

EAL was responsible for study design, review board's approval, data collection, statistical analysis, interpretation of results, and manuscript preparation. AR was responsible for study design and management, review board's approval, recruitment, data collection, interpretation of results, and manuscript preparation. CM was responsible for recruitment, data collection, and manuscript preparation. RBB and MDM were responsible for EMA, design and analysis, interpretation of results, and manuscript preparation. KJC was responsible for study design, InspireD app development, and manuscript preparation. BB and FF were responsible for study design, interpretation of results, and manuscript preparation. AG was responsible for InspireD app development, data collection, and manuscript preparation.

Conflicts of Interest

None declared.

References

1. Prince M, Knapp M, Guerchet M, McCrone P, Prina M, Comas-Herrera A, et al. Alzheimer's Society. 2014 Nov. 2014 Dementia UK: Update 2nd edn URL: https://www.alzheimers.org.uk/download/downloads/id/2323/dementia_uk_update.pdf [WebCite Cache ID 6zHEhzxXZ]
2. Lawrence V, Fossey J, Ballard C, Moniz-Cook E, Murray J. Improving quality of life for people with dementia in care homes: making psychosocial interventions work. Br J Psychiatry 2012 Nov;201(5):344-351 [FREE Full text] [doi: 10.1192/bjp.bp.111.101402] [Medline: 23118034]
3. Gonzalez J, Mayordomo T, Torres M, Sales A, Meléndez JC. Reminiscence and dementia: a therapeutic intervention. Int Psychogeriatr 2015 Oct;27(10):1731-1737. [doi: 10.1017/S1041610215000344] [Medline: 25765779]

4. Woods RT, Bruce E, Edwards RT, Elvish R, Hoare Z, Hounsome B, et al. REMCARE: reminiscence groups for people with dementia and their family caregivers - effectiveness and cost-effectiveness pragmatic multicentre randomised trial. Health Technol Assess 2012;16(48):v-xv, 1 [FREE Full text] [doi: 10.3310/hta16480] [Medline: 23211271]
5. Daly L, McCarron M, Higgins A, McCallion P. 'Sustaining Place' - a grounded theory of how informal carers of people with dementia manage alterations to relationships within their social worlds. J Clin Nurs 2013 Feb;22(3-4):501-512. [doi: 10.1111/jocn.12003] [Medline: 23170780]
6. Dassel KB, Carr DC, Vitaliano P. Does Caring for a Spouse With Dementia Accelerate Cognitive Decline? Findings From the Health and Retirement Study. Gerontologist 2017 Dec 01;57(2):319-328. [doi: 10.1093/geront/gnv148] [Medline: 26582383]
7. Hayes J, Zimmerman MK, Boylstein C. Responding to symptoms of Alzheimer's disease: husbands, wives, and the gendered dynamics of recognition and disclosure. Qual Health Res 2010 Aug;20(8):1101-1115. [doi: 10.1177/1049732310369559] [Medline: 20448273]
8. World Health Organization. 2017. Global action plan on the public health response to dementia 2017-2025 URL: http://apps.who.int/iris/bitstream/handle/10665/259615/9789241513487-eng.pdf;jsessionid=DAF085809E514B4306EDD110F4A0A0F6?sequence=1 [WebCite Cache ID 6zHEspu4I]
9. Gibson F. Reminiscence and life story work: A practice guide. London:: Jessica Kingsley Press; 2011.
10. Subramaniam P, Woods B. The impact of individual reminiscence therapy for people with dementia: systematic review. Expert Rev Neurother 2012 May;12(5):545-555. [doi: 10.1586/ern.12.35] [Medline: 22550983]
11. Karlsson E, Sävenstedt S, Axelsson K, Zingmark K. Stories about life narrated by people with Alzheimer's disease. J Adv Nurs 2014 Dec;70(12):2791-2799. [doi: 10.1111/jan.12429] [Medline: 24766325]
12. Lazar A, Thompson H, Demiris G. A systematic review of the use of technology for reminiscence therapy. Health Educ Behav 2014 Oct;41(1 Suppl):51S-61S [FREE Full text] [doi: 10.1177/1090198114537067] [Medline: 25274711]
13. Astell AJ, Ellis M, Bernardi L, Alm N, Dye R, Gowans G, et al. Using a touch screen computer to support relationships between people with dementia and caregivers. Interacting With Computers 2010;22(4):267-275.
14. Elfrink T, Juidema S, Kuinz M, Westerhof G. The effectiveness of creating an online life story book on persons with early dementia and their informal caregivers: a protocol of a randomized controlled trial. BMC Geriatr 2017;17 [FREE Full text] [Medline: 28431515]
15. Hamel AV, Sims TL, Klassen D, Havey T, Gaugler JE. Memory Matters: A Mixed-Methods Feasibility Study of a Mobile Aid to Stimulate Reminiscence in Individuals With Memory Loss. J Gerontol Nurs 2016 Jul 01;42(7):15-24 [FREE Full text] [doi: 10.3928/00989134-20160201-04] [Medline: 26870986]
16. Kerssens C, Kumar R, Adams AE, Knott CC, Matalenas L, Sanford JA, et al. Personalized technology to support older adults with and without cognitive impairment living at home. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):85-97. [doi: 10.1177/1533317514568338] [Medline: 25614507]
17. Mulvenna M, Doyle L, Wright T, Zheng H, Topping P, Boyle K, et al. Evaluation of card-based versus device-based reminiscing use photographic images. J Cyber Ther Rehab 2011;4(1):57-66 [FREE Full text]
18. Haesner M, Steinert A, O'Sullivan JL, Weichenberger M. Evaluating an Online Cognitive Training Platform for Older Adults: User Experience and Implementation Requirements. J Gerontol Nurs 2015 Aug;41(8):22-31; quiz 32. [doi: 10.3928/00989134-20150710-44] [Medline: 26248140]
19. Bleakley CM, Charles D, Porter-Armstrong A, McNeill MDJ, McDonough SM, McCormack B. Gaming for health: a systematic review of the physical and cognitive effects of interactive computer games in older adults. J Appl Gerontol 2015 Apr;34(3):NP166-NP189. [doi: 10.1177/0733464812470747] [Medline: 24652863]
20. Delello JA, McWhorter RR. Reducing the Digital Divide: Connecting Older Adults to iPad Technology. J Appl Gerontol 2015 Jun 17. [doi: 10.1177/0733464815589985] [Medline: 26084479]
21. Regier NG, Gitlin LN. Dementia-related restlessness: relationship to characteristics of persons with dementia and family caregivers. Int J Geriatr Psychiatry 2018 Jan;33(1):185-192. [doi: 10.1002/gps.4705] [Medline: 28332736]
22. Farina N, Page TE, Daley S, Brown A, Bowling A, Basset T, et al. Factors associated with the quality of life of family carers of people with dementia: A systematic review. Alzheimers Dement 2017 May;13(5):572-581 [FREE Full text] [doi: 10.1016/j.jalz.2016.12.010] [Medline: 28167069]
23. McDonnell E, Ryan AA. The experience of sons caring for a parent with dementia. Dementia (London) 2014 Nov;13(6):788-802. [doi: 10.1177/1471301213485374] [Medline: 24339083]
24. Melunsky N, Crellin N, Dudzinski E, Orrell M, Wenborn J, Poland F, et al. The experience of family carers attending a joint reminiscence group with people with dementia: A thematic analysis. Dementia 2015;14(6):842-859.
25. McHugh JE, Wherton JP, Prendergast DK, Lawlor BA. Teleconferencing as a source of social support for older spousal caregivers: initial explorations and recommendations for future research. Am J Alzheimers Dis Other Demen 2012 Sep;27(6):381-387. [doi: 10.1177/1533317512453491] [Medline: 22871904]
26. Archbold PG, Stewart BJ, Greenlick MR, Harvath T. Mutuality and preparedness as predictors of caregiver role strain. Res Nurs Health 1990 Dec;13(6):375-384. [Medline: 2270302]

27. Bech P, Olsen LR, Kjoller M, Rasmussen NK. Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 Mental Health subscale and the WHO-Five Well-Being Scale. Int J Methods Psychiatr Res 2003;12(2):85-91. [Medline: 12830302]
28. Spruytte N, Van AC, Lammertyn F, Storms G. The quality of the caregiving relationship in informal care for older adults with dementia and chronic psychiatric patients. Psychol Psychother 2002 Sep;75(Pt 3):295-311. [Medline: 12396755]
29. Gibson A, McCauley C, Mulvenna M, Ryan A, Laird EA, Bunting B, et al. Assessing usability testing for people living with dementia. In: Rehab.: Rehab; 2016 Presented at: Proceedings of the 4th Workshop on ICTs for improving Patients Rehabilitation Research Techniques; 2016; Lisbon p. 25-31 URL: https://kevincurran.org/papers/Accessing%20Usability%20Testing%20for%20People%20with%20Dementia%20-%20REHAB%202016.pdf [doi: 10.1145/3051488.3051492]
30. Aydin M, Harmsen F, van SK, Stegwee R. On the Adaptation of an agile Information systems development method. J Database Management 2005;16(4):24-40.
31. Mulvenna M, Gibson A, McCauley CO, Ryan AA, Bond R, Laird EA, et al. Behavioural Usage Analysis of a Reminiscing App for People Living with Dementia and their Carers. In: Proceedings of the European Conference on Cognitive Ergonomics. 2017 Presented at: ECCE 2017; 2017; Umea, Sweden p. 35-38. [doi: 10.1145/3121283.3121289]
32. Lewin K. A dynamic theory of personality. New York: McGraw-Hill; 1935.
33. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders. J Med Internet Res 2017 Mar 16;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
34. Intille S, Rondoni J, Kukla C, Ancona I, Bao L. A context-aware experience sampling tool. In: Proceedings of CHI 2003. 2003 Presented at: CHI 2003; 2003; Lauderdale, Florida p. 972-973.
35. Mulder I, Ter HG, Kort J. SocioXensor: Measuring user behaviour and user experience in context with mobile devices,. In: Proceedings of Measuring Behavior. 2005 Presented at: The 5th International Conference on Methods and Techniques in Behavioral Research; 2005; Wageningen, The Netherlands p. 355-358.
36. Wiebe DJ, Nance ML, Houseknecht E, Grady MF, Otto N, Sandsmark DK, et al. Ecologic Momentary Assessment to Accomplish Real-Time Capture of Symptom Progression and the Physical and Cognitive Activities of Patients Daily Following Concussion. JAMA Pediatr 2016 Nov 01;170(11):1108-1110. [doi: 10.1001/jamapediatrics.2016.1979] [Medline: 27617669]
37. Kingston D, Janes-Kelley S, Tyrrell J, Clark L, Hamza D, Holmes P, et al. An integrated web-based mental health intervention of assessment-referral-care to reduce stress, anxiety, and depression in hospitalized pregnant women with medically high-risk pregnancies: a feasibility study protocol of hospital-based implementation. JMIR Res Protoc 2015 Jan 16;4(1):e9 [FREE Full text] [doi: 10.2196/resprot.4037] [Medline: 25595167]
38. Cook N, Winkler SL. Acceptance, Usability and Health Applications of Virtual Worlds by Older Adults: A Feasibility Study. JMIR Res Protoc 2016 Jun 02;5(2):e81 [FREE Full text] [doi: 10.2196/resprot.5423] [Medline: 27256457]
39. Toscos T, Daley C, Heral L, Doshi R, Chen Y, Eckert GJ, et al. Impact of electronic personal health record use on engagement and intermediate health outcomes among cardiac patients: a quasi-experimental study. J Am Med Inform Assoc 2016 Jan;23(1):119-128. [doi: 10.1093/jamia/ocv164] [Medline: 26912538]
40. Webster JD, Bohlmeijer ET, Westerhof GJ. Mapping the Future of Reminiscence: A Conceptual Guide for Research and Practice. Res Aging 2010 May 24;32(4):527-564. [doi: 10.1177/0164027510364122]
41. Kordelaar KV, Vlak A, Kuin Y, Westerhof G. Een handleiding en werkboek voor intergenerationele gesprekken [Time Traveling: A manual for intergenerational groups]. In: Nijmegen: Centre for Psychogerontology. Radboud Universiteit Nijmegen: Centrum voor Psychogerontologie; 2007:2007.
42. Des JDC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [Medline: 14998794]
43. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiology 2012;65(3):301-308. [doi: 10.1016/j.jclinepi.2011.07.011] [Medline: 22169081]
44. Lyons KS, Sayer AG, Archbold PG, Hornbrook MC, Stewart BJ. The enduring and contextual effects of physical health and depression on care-dyad mutuality. Res Nurs Health 2007 Feb;30(1):84-98. [doi: 10.1002/nur.20165] [Medline: 17243110]
45. Tetz KB, Archbold PG, Stewart BJ, Messecar D, Hornbrook MC, Lucas SA. How frail elders evaluate their caregiver's role enactment: a scale to measure affection, skill, and attentiveness. J Fam Nurs 2006 Aug;12(3):251-275. [doi: 10.1177/1074840706291554] [Medline: 16837694]
46. Schumacher KL, Stewart BJ, Archbold PG. Mutuality and preparedness moderate the effects of caregiving demand on cancer family caregiver outcomes. Nurs Res 2007;56(6):425-433. [doi: 10.1097/01.NNR.0000299852.75300.03] [Medline: 18004189]
47. World Health Organization. 1998. Wellbeing Measures in Primary Health Care/The Depcare Project URL: http://www.euro.who.int/__data/assets/pdf_file/0016/130750/E60246.pdf [WebCite Cache ID 6yihIEpLT]

48. Woods RT, Orrell M, Bruce E, Edwards RT, Hoare Z, Hounsome B, et al. REMCARE: Pragmatic Multi-Centre Randomised Trial of Reminiscence Groups for People with Dementia and their Family Carers: Effectiveness and Economic Analysis. PLoS One 2016;11(4):e0152843 [FREE Full text] [doi: 10.1371/journal.pone.0152843] [Medline: 27093052]
49. Henkel V, Mergl R, Kohnen R, Allgaier A, Möller H, Hegerl U. Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. Gen Hosp Psychiatry 2004;26(3):190-198. [doi: 10.1016/j.genhosppsych.2004.02.003] [Medline: 15121347]
50. Liwowsky I, Kramer D, Mergl R, Bramesfeld A, Allgaier A, Pöppel E, et al. Screening for depression in the older long-term unemployed. Soc Psychiatry Psychiatr Epidemiol 2009 Aug;44(8):622-627. [doi: 10.1007/s00127-008-0478-y] [Medline: 19048174]
51. de WM, Pouwer F, Gemke RJBJ, Delemarre-van DWHA, Snoek FJ. Validation of the WHO-5 Well-Being Index in adolescents with type 1 diabetes. Diabetes Care 2007 Aug;30(8):2003-2006. [doi: 10.2337/dc07-0447] [Medline: 17475940]
52. Löwe B, Spitzer RL, Gräfe K, Kroenke K, Quenter A, Zipfel S, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. J Affect Disord 2004 Feb;78(2):131-140. [Medline: 14706723]
53. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
54. IBM. IBM SPSS Statistics for Windows, Version 23.0. Armonk, New York: IBM Corporation; 2016.
55. Bannon W. Missing data within a quantitative research study: How to assess it, treat it, and why you should care. J Am Assoc Nurse Pract 2015 Apr;27(4):230-232. [doi: 10.1002/2327-6924.12208] [Medline: 25676704]
56. Pallant J. SPSS Survival Manual. New York: McGraw Hill; 2010.
57. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]
58. Sheikh J, Yesavage J. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clin Gerontol 1986;5(1/2):165-173 Published online: 25 Oct 2008.
59. Kerkhof YJF, Graff MJL, Bergsma A, de VHHM, Dröes R. Better self-management and meaningful activities thanks to tablets? Development of a person-centered program to support people with mild dementia and their carers through use of hand-held touch screen devices. Int Psychogeriatr 2016 Dec;28(11):1917-1929. [doi: 10.1017/S1041610216001071] [Medline: 27425002]
60. 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]
61. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
62. Hale K, Capra S, Bauer J. A Framework to Assist Health Professionals in Recommending High-Quality Apps for Supporting Chronic Disease Self-Management: Illustrative Assessment of Type 2 Diabetes Apps. JMIR Mhealth Uhealth 2015;3(3):e87 [FREE Full text] [doi: 10.2196/mhealth.4532] [Medline: 26369346]
63. Alzheimer's Research UK. Women and dementia: A marginalised majority. Cambridge: Alzheimer's Research UK; 2015. URL: https://www.alzheimersresearchuk.org/wp-content/uploads/2015/03/Women-and-Dementia-A-Marginalised-Majority1.pdf [accessed 2018-08-08] [WebCite Cache ID 71rtBEgrl]

Abbreviations

ANOVA: analysis of variance
EMA: ecological momentary assessment
ICT: information and communication technology
InspireD: Individual-Specific Reminiscence for People living with Dementia
IT: information technology
QCPR: Quality of Carer-Patient Relationship scale
RCT: randomized controlled trial
UI: user interface
WHO: World Health Organization
WHO-5: World Health Organization–Five Well-Being Index

Edited by G Eysenbach; submitted 19.12.17; peer-reviewed by ME Kamasak, Q Yuan, R El Shafie, K Hale; comments to author 18.03.18; revised version received 10.05.18; accepted 18.06.18; published 11.09.18.

Please cite as:
Laird EA, Ryan A, McCauley C, Bond RB, Mulvenna MD, Curran KJ, Bunting B, Ferry F, Gibson A
Using Mobile Technology to Provide Personalized Reminiscence for People Living With Dementia and Their Carers: Appraisal of Outcomes From a Quasi-Experimental Study
JMIR Ment Health 2018;5(3):e57
URL: http://mental.jmir.org/2018/3/e57/
doi:10.2196/mental.9684
PMID:30206053

Original Paper

Differences in the Use and Opinions About New eHealth Technologies Among Patients With Psychosis: Structured Questionnaire

Lucia Bonet[1]; Blanca Llácer[2], PhD; Miguel Hernandez-Viadel[3], PhD, MD; David Arce[4], Research Engineer; Ignacio Blanquer[4], PhD (ComSci); Carlos Cañete[3], PhD, MD; Maria Escartí[5], PhD, MD; Ana M González-Pinto[6,7], PhD, MD; Julio Sanjuán[1,2,3], PhD, MD

[1]Department of Clinic Medicine, School of Medicine, University of Valencia, Valencia, Spain

[2]Centre of Biomedical Investigation in Mental Health (CIBERSAM), Spanish Government Carlos III Health Institute, Valencia, Spain

[3]Department of Mental Health, Sanitary Research Institute of Valencia (INCLIVA), Hospital Clínico of Valencia, Valencia, Spain

[4]Institute of Instrumentation for Molecular Imaging (I3M), Joint centre CSIC & Universitat Politècnica de València, Valencia, Spain

[5]Department of Mental Health, Font of Sant Lluís Primary Care Centre, Valencia, Spain

[6]Centre of Biomedical Investigation in Mental Health (CIBERSAM), Spanish Government Carlos III Health Institute, Vitoria, Spain

[7]Department of Psychiatry, BioAraba Research Institute, Araba University Hospital, University of the Basque Country (EHU/UPV), Vitoria, Spain

Corresponding Author:
Lucia Bonet
Department of Clinic Medicine
School of Medicine
University of Valencia
Medicine Department, 2nd Floor
15 Blasco Ibañez Avenue
Valencia, 46010
Spain
Phone: 34 600745550
Email: lbonetm@gmail.com

Abstract

Background: Despite a growing interest in the use of technology in order to support the treatment of psychotic disorders, limited knowledge exists about the viability and acceptability of these eHealth interventions in relation to the clinical characteristics of patients.

Objective: The objective of this study was to assess the access and use of, as well as experiences and interest in, new technologies using a survey of patients diagnosed with early psychosis compared with a survey of patients diagnosed with chronic psychotic disorders.

Methods: We designed a structured questionnaire. This questionnaire was divided into five parts: (1) clinical and demographic information, (2) access and use of the internet, (3) use of the internet in relation to mental health, (4) experiences with technology, and (5) patients' interest in eHealth services. In total, 105 patients were recruited from early psychosis units (n=65) and recovery units (n=40).

Results: In this study, 84.8% (89/105) of the patients had access to the internet and 88.6% (93/105) owned an electronic internet device. In total, 71.3% (57/80) of patients who owned a mobile phone were interested in eHealth systems and 38.2% (37/97) reported negative experiences related to the internet usage. We observed differences between the groups in terms of device ownership (P=.02), the frequency of internet access (P<.001), the use of social media (P=.01), and seeking health information (P=.04); the differences were found to be higher in the early psychosis group. No differences were found between the groups in terms of the use of internet in relation to mental health, experiences and opinions about the internet, or interest in eHealth interventions (P=.43).

Conclusions: The availability and use of technology for the participants in our survey were equivalent to those for the general population. The differences found between the groups in relation to the access or use of technology seemed to due to age-related factors. The use of technology involving mental health and the interest in eHealth interventions were mainly positive and equivalent

between the groups. Accordingly, this group of patients is a potential target for the emerging eHealth interventions, regardless of their clinical status. However, 28.7% (23/80) of the studied patients rejected the use of internet interventions and 38.2% (37/97) had unpleasant experiences related to its usage; thus, more in-depth studies are needed to better define the profile of patients with psychosis who may benefit from eHealth treatments.

(JMIR Ment Health 2018;5(3):e51) doi:10.2196/mental.9950

KEYWORDS

eHealth; internet; mobile phone; viability; acceptability; psychosis; schizophrenia

Introduction

The relevance of early intervention (EI) in psychotic disorders in order to prevent the pathological development of the illness is well known [1]. However, some studies have shown that the current models of EI do not produce any different results in terms of efficacy or efficiency when compared with treatment as usual [2,3]. In this vein, technological developments could make a difference by adapting these traditional models of psychiatric and psychological health care to an electronic form, which would allow interactive and more personalized tracking of patients and Web-delivered therapy such as psychoeducational services or cognitive behavioral treatments [4]. These technological health interventions are known as eHealth [5]. The recent examples of these interventions that are being currently tested are Actissist [6], Prime [7], and SlowMo [8].

Nevertheless, before proceeding further in developing these eHealth interventions, it is important to better understand the relationship between patients with psychosis and technology resources. Psychotic disorders are characterized by their clinical heterogeneity [9]; thus, it is necessary to study if these eHealth interventions are equally accepted for all patients with psychosis, regardless of their demographics or clinical characteristics, especially if they are in an early psychosis (EP) condition or a chronic psychosis (CP) condition.

First, it is important to assess whether the access and use of technology are equivalent between EP and CP patients and whether the access and use are equivalent to those among the general population. Depp et al [10] conducted a survey of CP patients and found that these patients had substantial cognitive and functional deficits and that high punctuations in these impairments were related to the less use of technology. Moreover, in 2014, the National Alliance on Mental Illness (NAMI) [11] showed that 54% of American patients with schizophrenia owned a mobile phone compared with 64% of the general American population [12]; similar results have been shown in other studies [13]. However, recent studies have shown that these rates have changed and that the access of these patients to technology is similar to that of the general population at the moment [14-16].

Second, 80% of patients with psychosis are permitted to use internet resources in relation to their illness management [17]. Nevertheless, we could not find any study that investigated whether this use of technology is equivalent between EP and CP patients, who are usually more aged persons with more associated morbidities [10].

Third, despite the majority of patients who report positive feelings and experiences in response to the internet usage [4,18], there are some patients who experience anxiety or paranoid feelings while using this resource [18]. Moreover, some patients admitted that they had stopped taking medication on their own because of the information they read on the internet [17]. In relation to this, it is important to better understand the effect that technologies have on these patients and whether these experiences are similar between EP and CP patients.

Finally, there are several studies that have confirmed the interest of patients experiencing psychotic disorders in using the emerging eHealth systems to help them cope with their illness [4,14,18,19]. Specially, it has been found that 60%-75% of patients with psychosis would be interested in receiving information and feedback from their clinicians [19,20] and in contacting them in case of emergency [20]. However, there is a lack of studies that have assessed this interest in relation to the evolution of the disease (EP compared with CP). There are a few studies that have found some controversial results when studying this interest among individuals of different age groups. Some of these studies have suggested that younger patients would be more willing to endorse eHealth treatments [14,18], while others have suggested the opposite [21,22]. Consequently, it is necessary to study the variations in the interest in these services in relation to the evolution of the illness.

The main objective of this study was to assess the access and use of and experiences with technology in a survey of patients diagnosed with EP compared with a survey of patients diagnosed with CP disorder. In addition, we aimed to analyze the interest in these two groups regarding using an eHealth system and regarding the different tracking eHealth services suggested.

Methods

Measures and Design

The data were collected through a cross-sectional questionnaire that we designed for the purpose of this investigation. To elaborate this questionnaire, we reviewed studies about the use, access, and impact of technology on patients with psychosis. Based on these studies, we elaborated the survey, which is divided into five parts: the items for the first part, which aims to assess clinical and demographic information, and the items for the second part, which measures the access and use of the internet, mobile, and social media, were taken from the Spanish National Statistics Institute [23] survey and from studies by Trefflich et al [17] and Robotham et al [24]. In addition, the items for the third part of the questionnaire, which assesses the use of internet in relation to mental health, and the items for the

fourth part, which measures experiences with technology and the effect of internet usage on patients' health, were based on a survey of the NAMI [11] and on studies by Gay et al [18], Miller et al [25], and Borzekowski et al [26]. The last part of the survey, which rates the interest of the patients in using an eHealth app and their interest in different tracking and reminder services, was an originally developed section.

Once the instrument was made and prior to its use, a pilot study was conducted to check the acceptability and relevance of the measure. Overall, 14 representative patients participated; consequently, 3 ambiguous items were corrected in order to make them easier to understand for the patients, and 2 redundant items were removed.

The 10-minute, structured questionnaire (Multimedia Appendix 1) was completed face-to-face. Initially, the patients were informed about data extraction ethics and confidentiality following the information sheet (Multimedia Appendix 2); subsequently, the patients completed the questionnaire. All the patients signed the informed consent before participating in this survey. The survey was conducted from February to May 2017 and was approved by the Clinical Hospital of Valencia's Ethics Committee.

Sample and Recruitment

A total of 113 participants were eligible for inclusion. They met the following criteria: (1) diagnosis of a psychotic disorder according to the International Classification of Diseases, Tenth Revision [27]; (2) clinically stable; (3) outpatient from the first episode units at the Clinical Hospital of Valencia and from the Primary Care Centre Font of Sant Lluis in Valencia or outpatient from association for comprehensive care of the mental health patient or from aid association for mental health patients in the Valencia community recovery units; and (4) able to communicate in Spanish. Patients were excluded if they had severe cognitive impairments or did not complete the entire questionnaire.

Data Analysis

We analyzed data using the statistical program IBM SPSS Statistics version 22. We excluded 8 patients from this analysis for not having totally filled the survey; due to this, data of 105 patients were considered for the analysis. The cohort was divided into two groups: the EP group, with a duration of illness of ≤72 months, and the CP group, with a duration of illness >72 months. This division was based on the fact that EP patients are treated in EP units until a maximum period of 72 months. Descriptive statistics (mean, standard deviation, frequencies, and percentages) were determined, and chi-square test and analysis of variance were performed in order to compare the differences between the EP and CP groups.

Results

The data in Tables 1, 2, 3, and 4 are shown in the following order: First, the EP results are shown, followed by the CP results and the total results (which are the global results of the sample in each category). It is important to mention that the sample is not the same in every category due to the fact that some questions in the survey were exclusionary. If the patients did not fulfill the profile for one question, they did not have to complete the rest of the questions that were related to the first one. We have marked this condition in every table.

Sample Characteristics

A total of 105 participants were enrolled in the study. Based on the duration of their illness, we assigned 65 patients to the EP group (≤72 months) and 40 patients to the CP group (>72 months). The mean age of the sample was 38.1 (SD 13) years; the patients were mostly male (76/105, 72.4%) and single (89/105, 84.8%) and had achieved a secondary level of education (compulsory schooling: 26/105, 24.8%; secondary education: 39/105, 37.1%).

We found significant differences between the two groups. EP patients were mostly in the first episode of psychosis (FEP), while CP patients were mostly diagnosed with schizophrenia. The duration (months) of illness was higher in the CP group. The patients in the EP group were younger and mostly employed, while those in the CP group were mostly unable to work or were not employed. There were no significant between-group differences in terms of gender, marital status, or the level of education. These clinical and sociodemographic characteristics are displayed in Table 1.

Access and Use of the Internet, Mobile and Social Media

Of all the participants, 84.8% (89/105) had access to the internet in the 3 months prior to the study, and there was high electronic device availability in the survey (93/105, 88.6%). After the first two questions, 8 patients did not continue with completing the survey as they were considered "electronic excluded" patients because they were not using or had not used the internet sufficiently to consider their experience relevant for the aim of this study. From that moment on, the total sample consisted of 97 patients (EP, n=63; CP, n=34).

Differences between the groups (Table 2) were found in terms of electronic device availability (χ^2_5=13.8, P=.02), the frequency of access to the internet (χ^2_2=31.8, P<.001), and the use of social media (χ^2_4=13.9, P=.01). Electronic device availability was higher in the EP group (63/65, 97%) than in the CP group (30/40, 75%), and while 81% (51/63) patients in the EP group had daily access to internet, 52.9% (8/34) of the patients in the CP group had only weekly access. However, no differences were observed in terms of the type of device used to access (χ^2_2=5.6, P=.06), mobile ownership (χ^2_5=10.2, P=.07), or the most used functions of the mobile phone, which were calls (74/88, 84.1%; χ^2_1=0.7, P=.41) and texting or WhatsApp (72/88, 83.8%; χ^2_1=0.4, P=.51) for both groups. Social media ownership was higher in the EP group (51/63, 81%) than in the CP group (15/34, 44.1%); however, Facebook was the most used social media site in both the groups (47/66, 72.3%; χ^2_1=1.6, P=.21), and patients' main goal in using this social media platform was to communicate with people (55/66, 83.3%; χ^2_1=1.4, P=.24).

Internet and Mental Health

Internet is a resource that 61.9% (60/97) of the patients used to seek information about health. EP patients (45/63, 71.4%) used this resource to a greater extent than CP patients (15/34, 44.1%; χ^2_5=11.5, *P*=.04). The most wanted information was regarding symptoms (47/60, 78.3%) or diagnosis (40/60, 66.7%), which was more sought after by CP patients (14/15, 93.3%) than by EP patients (26/45, 57.8%; χ^2_1=6.4, *P*=.01). Of all the patients, 37.1% (36/97) stated that the internet was their first resource for seeking health information, whereas 58.8% (57/97) consulted clinical services as a first option.

In relation to the feelings that the use of internet provided to the patients, we found that 60.9% (59/97) felt socially linked when using internet and that 78.4% (76/97) felt informed. However, 22.7% (22/97) of the patients felt frustrated or anxious in relation to the internet and 19.6% (19/97) felt suspicious or paranoid.

Table 1. Demographic and clinical characteristics.

Characteristics	Early psychosis (N=65)	Chronic psychosis (N=40)	Total (N=105)	*P* value (χ^{2a} or t^b, df^c)
Diagnosis, n (%)				**<.001 (61.9[a], 7)**
Schizophrenia	9 (13.8)	29 (72.5)	38 (36.2)	
First episode of psychosis	44 (67.7)	0 (0.0)	44 (41.9)	
Other psychotic disorder[d]	12 (18.5)	18 (27.5)	23 (21.9)	
Duration of Illness (months), mean (SD)	28.8 (21.3)	253.3 (115)	114.3 (131.3)	<.001 (235.9[b], 1)
Age (years), mean (SD)	32.9 (11.8)	46.6 (10.3)	38.1 (13)	<.001 (−6.1[b], 103)
Gender, n (%)				**.38 (.8[a], 1)**
Female	16 (24.6)	13 (32.5)	29 (27.6)	
Male	49 (75.4)	27 (67.5)	76 (72.4)	
Marital status, n (%)				**.07 (7.2[a], 3)**
Single	56 (86.2)	33 (82.5)	89 (84.8)	
Married	7 (10.8)	1 (2.50)	8 (7.6)	
Widowed	0 (0)	1 (2.50)	1 (1.0)	
Divorced	2 (3.1)	5 (12.5)	7 (6.7)	
Education, n (%)				**.43 (3.8[a], 4)**
Primary school	11 (16.9)	8 (20)	19 (18.1)	
Compulsory schooling[e]	17 (26.2)	9 (22.5)	26 (24.8)	
Secondary education	22 (33.8)	17 (42.5)	39 (37.1)	
University degree	15 (23.1)	6 (15.0)	21 (20.0)	
Employment status, n (%)				**<.001 (27.7[a], 6)**
Employed	18 (27.7)	4 (10.0)	22 (21.0)	
Not employed	16 (24.6)	11 (27.5)	27 (25.7)	
Student	16 (24.6)	1 (2.5)	17 (16.2)	
Unable to work	9 (13.8)	18 (45.0)	27 (25.7)	
Others	6 (9.3)	6 (15.0)	12 (11.5)	

[a]Chi-square (χ^2) values.

[b]Student *t* values.

[c]*df*: degrees of freedom.

[d]Referring more than one psychotic episode or a specific disorder (bipolar, schizophreniform, schizoaffective, major depression, personality disorder).

[e]Until the age of 16 years.

Table 2. Access to and use of the internet, mobile, and social media.

Access and use of technology	Early psychosis (N=65), n (%)	Chronic psychosis (N=40), n (%)	Total (N=105), n (%)	*P* value (χ^2, *df*[a])
Internet access (last 3 months)	**N=65**	**N=40**	**N=105**	**.05 (7.5, 3)**
Yes	59 (90.8)	30 (75)	89 (84.8)	
No	6 (9.2)	10 (25)	16 (15.2)	
Electronic device availability	**N=65**	**N=40**	**N=105**	**.02 (13.8, 5)**
Yes	63 (97)	30 (75)	93 (88.6)	
No	2 (3)	10 (25)	7 (11.4)	
Device type	**N=63**	**N=30**	**N=93**	**.06 (5.6, 2)**
Computer	16 (25.4)	13 (43.3)	29 (31.2)	
Mobile	47 (74.6)	16 (53.3)	63 (67.7)	
Tablet	0 (0)	1 (3.3)	1 (1.1)	
Frequency of internet access	**N=63**	**N=34**	**N=97**	**lt;.001 (31.8, 2)**
Daily	51 (81)	11 (32.4)	62 (63.9)	
Weekly	3 (4.8)	18 (52.9)	21 (21.6)	
Less than once a week	9 (14.3)	5 (14.7)	14 (14.4)	
Mobile ownership	**N=63**	**N=34**	**N=97**	**.07 (10.2, 5)**
Yes, cell phone	7 (11.1)	6 (17.6)	13 (13.4)	
Yes, mobile phone	54 (85.7)	21 (61.8)	75 (77.3)	
No	2 (3.2)	7 (20.6)	9 (9.3)	
Mobile use[b]	**N=61**	**N=27**	**N=88**	
Calls	50 (82)	24 (88.9)	74 (84.1)	.41 (.7, 1)
Texting or WhatsApp	51 (83.6)	21 (77.8)	72 (83.8)	.51 (.4, 1)
Social media ownership	**N=63**	**N=34**	**N=97**	**.01 (13.9, 4)**
Yes	51 (81)	15 (44.1)	66 (68)	
No	12 (19)	19 (55.9)	31 (32)	
Social media site[b]	**N=51**	**N=15**	**N=66**	
Facebook	351 (68.6)	12 (80)	47 (72.3)	.21 (1.6, 1)
WhatsApp groups	34 (66.7)	12 (80)	46 (69.7)	.32 (.9, 1)
Social media use[b]	**N=51**	**N=15**	**N=66**	
To communicate with people	44 (86.3)	11 (73.3)	55 (83.3)	.24 (1.4, 1)
To stay informed	35 (68.6)	9 (64.3)	44 (67.7)	.76 (.1, 1)

[a]*df*: degrees of freedom.

[b]Sample reduction because of a previous exclusionary question.

Table 3. Internet and mental health.

Experiences and opinions about internet	Early psychosis (N=63), n (%)	Chronic psychosis (N=34), n (%)	Total (N=97), n (%)	*P* value (χ^2, *df*[a])
Internet used to seek health information	**N=63**	**N=34**	**N=97**	**.04 (11.5, 5)**
Yes	45 (71)	15 (44)	60 (62)	
No	18 (29)	19 (56)	37 (38)	
Most sought after health information[b]	**N=45**	**N=15**	**N=60**	
Symptoms	36 (80)	11 (73)	47 (78)	.59 (.3, 1)
Diagnosis	26 (58)	14 (93)	40 (67)	.01 (6.4, 1)
Internet: first information resource	**N=63**	**N=34**	**N=97**	.13 (5.7, 3)
Agree	27 (43)	9 (27)	36 (37)	
Disagree	36 (57)	25 (74)	61 (63)	
Agreement on internet feelings[c]	**N=63**	**N=34**	**N=97**	
Socially linked	28 (60)	21 (62)	59 (61)	.93 (.9, 5)
Informed	53 (84)	23 (68)	76 (78)	.24 (5.5, 4)
Frustrated or Anxious	11 (18)	11 (32)	22 (23)	.08 (8.3, 4)
Suspicious or Paranoid	13 (21)	6 (18)	19 (20)	.46 (3.6, 4)
Agreement on internet experiences[c]	**N=63**	**N=34**	**N=97**	
Internet as a benefit for mental health	27 (43)	18 (53)	45 (46)	.23 (5.6, 4)
Unpleasant experiences related to internet usage	25 (40)	12 (35)	37 (38)	.92 (.9, 4)
Stopped taking medication because of internet information	4 (6)	4 (12)	8 (8)	.84 (1.4, 4)
Relapse related to internet usage	20 (32)	4 (12)	24 (25)	.17 (6.4, 4)
Excessive time spent on internet	20 (32)	6 (18)	26 (27)	.25 (5.4, 4)
Internet increases social isolation	12 (19)	5 (15)	17 (16)	.69 (2.3, 4)

[a]*df*: degrees of freedom.

[b]Sample reduction because of a previous exclusionary question.

[c]Sum of individual scores of "Strongly agreed" and "Somewhat agreed" in each factor.

Regarding experiences related to internet usage, we found that 46.4% (45/97) of the patients thought that the internet is beneficial to their mental health, while 38.2% (37/97) had unpleasant experiences related to its usage, and 24.8% (24/97) patients had experienced relapses perceived as directly related to internet usage. Moreover, 8.3% (8/97) patients had stopped taking medication on their own because of the information they read on the internet. Excessive time on the internet was a concern for 26.8% (26/97) of the patients and 16.2% (17/97) thought that internet increases social isolation. As displayed in Table 3, we could not find any significant between-group differences in terms of the feelings about the internet or experiences related to its usage.

Interest in eHealth Systems (Mobile Phone App)

This part of the survey was completed only by patients who owned a mobile phone. For this reason, the sample size was reduced to 80 patients (EP, n=59; CP, n=21). Of all the patients, 71.3% (57/80) were interested in owning an eHealth app, with no significant differences observed between the EP and CP groups (χ^2_4=3.9; *P*=.43); furthermore, no significant differences were observed in terms of age of the sample (F_1=.08, *P*=.93). The reason for not being interested was "I do not think I will benefit from it" (14/23, 60.9%) or "I have enough information" (6/23, 26.1%).

The services that were perceived as the most interesting were as follows: clinician contact alarm (60/80, 75.1%) and a reminder for clinical appointments (58/80, 72.6%). Mood, mental health, and side effect tracking were perceived as equally interesting (51/80, 63.8%), while the least interesting function for the patients was the reminder to take medication (41/80, 51.3%). As shown in Table 4, no significant differences were found between the groups in terms of their interest in any of the services suggested. Furthermore, no significant differences were found regarding the age of the sample and interest in mood and mental health service (F_1=1.31, *P*=.27), interest in side effect tracking (F_1=1.44, *P*=.24), reminder for clinical appointments (F_1=.99, *P*=.37), reminder to take medication (F_1=.2.35, *P*=.11), and clinician contact alarm (F_1=.47, *P*=.63).

Table 4. Interest in eHealth systems (app).

Opinions about eHealth app services	Early psychosis (N=59), n (%)	Chronic psychosis (N=21), n (%)	Total (N=80), n (%)	*P* value (χ^2, *df*[a])
App interest				**.43 (3.9, 4)**
Yes	42 (71.2)	15 (71.4)	57 (71.3)	
No	**17 (28.8)**	**6 (28.6)**	**23 (28.7)**	
I do not think I will benefit from it[b]	11 (64.7)[c]	3 (50)[d]	14 (60.9)[e]	
I have enough information[b]	5 (29.4)[c]	1 (16.7)[d]	6 (26.1)[e]	
Others[b]	1 (5.9)[c]	2 (33.3)[d]	3 (13)[e]	
App services				
Mood and mental health tracking				**.77 (1.8, 4)**
Interested[f]	38 (64.4)	13 (61.9)	51 (63.8)	
Indifferent	5 (8.5)	3 (14.3)	8 (10)	
Not interested[g]	16 (27.1)	5 (23.8)	21 (26.3)	
Side effect tracking				**.39 (4.1, 4)**
Interested[f]	40 (67.8)	11 (52.3)	51 (63.8)	
Indifferent	4 (6.8)	4 (19)	8 (10)	
Not interested[g]	15 (25.4)	6 (28.5)	21 (26.3)	
Reminder of clinical appointments				**.82 (1.5, 4)**
Interested[f]	41 (69.5)	17 (80.9)	58 (72.6)	
Indifferent	4 (6.8)	1 (4.8)	5 (6.3)	
Not interested[g]	14 (23.7)	3 (14.3)	17 (21.3)	
Reminder to take medication				**.32 (4.7, 4)**
Interested[f]	29 (49.1)	10 (57.1)	41 (51.3)	
Indifferent	11 (18.6)	2 (9.5)	13 (16.3)	
Not interested[g]	19 (32.2)	7 (33.3)	26 (32.6)	
Clinician contact alarm				**.12 (7.3,4)**
Interested[f]	44 (74.5)	16 (76.2)	60 (75.1)	
Indifferent	4 (6.8)	—	4 (5)	
Not interested[g]	11 (18.6)	5 (23.8)	16 (20)	

[a]*df*: degrees of freedom.

[b]Sample reduction because of a previous exclusionary question.

[c]N=17.

[d]N=6.

[e]N=23.

[f]Sum of individual scores of “Very interested” and “Somewhat interested” in each factor.

[g]Sum of individual scores of “Not very interested” and “Not at all interested” in each factor.

Discussion

Access and Use of Technology

The rates of accessibility and usability of the internet, mobile, and social media in our surveyed sample were high and very similar to the rates we found in the general Spanish population [23]. These results contradict the lower rates obtained by the NAMI study in 2014 [11] and are more similar to the results of recent studies [14-16,18,28], which found that the access and use of technology in patients diagnosed with psychotic disorders are equivalent to those in the general population. The differences between the two comparison groups in this study suggested that the access and use are not equivalent between EP and CP patients. As we found, CP patients had less electronic device

availability (CP: 30/40, 75%; EP: 63/65, 97%) as well as lower rates of daily access to the internet (CP: 11/34, 32.4%; EP: 51/63, 81%) and use of social media (CP: 15/34, 44.1%; FEP: 51/63, 81%) than EP patients. However, these differences were not only found in previous studies on patients with psychosis [14,17,24,29] but also found in studies on the general Spanish population [23,30]. All these studies agreed that younger patients (18-34 years) have the highest rates of access and use of technology and that these rates start to decrease with the increasing age. In relation to this, we suggest that the differences found between EP and CP patients might be more related to the fact that EP patients were younger than CP patients (P<.001) than to a pathologically related issue.

Use of the Internet Related to Mental Health

In accordance with previous studies [17,23,28,31], the internet is a resource that both patients and the general population use in order to seek information about health. Moreover, nearly 40% (39/97) of our patients admitted that the internet is their first source of health information. However, in accordance with previous studies [17], EP patients used this resource to a greater extent than CP patients (EP: 45/63, 71.4%; CP: 15/34, 44.1%; P=.04). Nevertheless, it is important to note that nearly 56% (19/36) of CP patients and 29% (18/63) of EP patients did not use the internet to seek health information and that nearly 63% (61/97) of patients did not regard internet as their first source of information. These results suggest that despite the fact that the internet is an accessible and quick resource to obtain information [28,32], patients still rely on clinicians as their first source of health information.

Experiences and Opinions About the Internet

In line with previous studies [4,18,32], between 60.9% (59/97) and 78.4% (76/97)of patients reported positive experiences related to the internet usage. However, 22.7% (22/97) of the patients felt frustrated or anxious in relation to the internet, and 19.6% (19/97) felt suspicious or paranoid. Moreover, 38.2% (37/97) of the patients had had unpleasant experiences related to internet usage, 24.8% (24/97) had experienced relapses perceived as directly related to its usage, and 8.3% (8/97) of the sample had stopped taking their medication on their own decision because of the information that was read on the internet. It should be noted that despite the fact that the access and use of technology were found to be higher in the EP group, there were no between-group differences in relation to their experiences of or opinions on internet usage. However, these negative experiences have been found in previous studies [18,29,32], and they suggest that although internet could be a great resource to improve the empowerment of the patients in the management of their illness [32] or as an entertainment resource [18], it could also be a source of stress by causing anxious or paranoid feelings [18,29,32]. It is important to mention that 50%-56% of the general Spanish population agrees with “being worried about internet, social media, and government use of personal information given on the internet” [33]; in accordance with this, we suggest that new technologies are a source of information that could be interpreted as a false alarm signal that may trigger paranoid symptoms. However, we could not find any studies concerning this issue.

Moreover, although 60.9% (59/97) of the patients felt socially linked when using the internet, 26.8% (26/97) admitted to spending excessive time on it and 16.2% (17/97) thought that internet increases social isolation. This enhancement of social isolation has also been reported in studies in the general population [34]. In accordance with previous studies, social isolation is a risk factor for psychosis [35], and it is one of the key relapse factors following the FEP [36].

Interest in eHealth Systems (Mobile Phone App)

Consistent with previous studies [4,14,19], the interest in owning an eHealth system (mobile phone app) in our sample was high (57/80, 71.3%), with no differences observed between the two comparison groups. Moreover, we could not find any significant differences between the groups in terms of their interest in the different eHealth services suggested or when comparing the age of the sample. This result has been found in a systematic review of previous acceptance studies [19], which concluded that there is no difference between clinical and demographic characteristics and the acceptance of eHealth interventions. In line with this, the high acceptance of eHealth interventions in our sample could be regarded as a potential confirmation that patients with psychotic disorders are a good target for these emerging interventions, with no differences related to the length of the illness.

However, although the differences were not statistically significant, on comparing both groups, we found that the percentages of interest were higher in the CP group than in the EP group regarding “reminder services” (clinical appointments and taking medication). In a previous study [21], it was found that the older the patients were, the more reminders they would select. In line with that study, we suggest that CP patients, being more aged and impaired than EP patients, as shown in Table 1 and in previous studies [10], could regard reminder services as a helpful tool to manage their illness, whereas the EP group, being younger and having better social support and less associated impairments, would not regard this service as useful.

On the other hand, EP patients found the “tracking services” (mood, mental health and, side effects) more interesting than CP patients. In a systematic review of previous publications [19], it was found that the interest of patients in receiving psychoeducative and symptom information increased to 90% in the EP sample. According to this finding, we speculated that the EP group would consider “tracking services” more interesting due to their more recent diagnosis and need to better understand their illness, whereas the more experienced CP patients would not consider this service useful.

However, as noted before, there were no significant differences between the groups; thus, initially, patients in both the groups (EP and CP) would be interested in any service in an equivalent way regardless of their age.

Finally, it is important to mention that the most interesting service for the patients was the “contact alarm to the clinicians in case of emergency” (60/80, 75.1%); the interest shown by both groups was nearly the same (EP: 44/59, 74.5%; CP: 16/21, 76.2%). This service must be a priority in eHealth developments. Patients are asking for more personalized, interactive, and closer

clinical attention [14,19,20], which could lead to a greater improvement in psychosis EI [4]. However, as noticed in previous studies [37,38], regarding the clinical implications associated to these interventions, it is highly important to design these feedback systems taking into consideration the clinicians' perspective to not overwhelm their capacities to respond to this systems.

Limitations

This study has some limitations. First, we cannot generalize the results to a broader population of individuals with psychotic disorders. We could not conduct a randomized selection of the sample; therefore, it was selected for the purpose of the aim of this study. Moreover, the small sample size (N=105) and the fact that 72.4% (46/105) of the patients were males with a mean age of 38.1 (SD 13) years caused our sample to not be representative of the demographic distribution of individuals with psychosis. In addition, some demographic information, such as the ethnicity of the sample was not collected. However, it is important to note that most of our results are consistent with those of previous publications; thus, we could infer that in a larger, randomized sample, the results would be similar to the ones we obtained in this study.

Second, the data were obtained from a questionnaire designed for the purpose of this investigation. Even though it was based on a previous review of publications and we conducted a pilot study to test its validity, our survey was not a standardized or a properly validated instrument for individuals with psychotic disorders. The quality of data obtained was affected for this reason. Moreover, most of the items in the survey measured nominal information, which hampered the performance of more complex statistical analyses. In relation to this, some items measured opinions or patients' perceptions, and we did not include an open text-box in order to better understand the responses given by the patients to these items.

Finally, regarding items of the final section of the questionnaire, since eHealth services are rapidly progressing, future updates of these items would be needed.

Implications and Orientations for Future Research

This study highlighted the viability and relatively high acceptability of eHealth interventions in a sample of patients diagnosed with psychotic disorders. However, some disregarded issues must guide future investigations in the area of eHealth and psychotic disorders.

First, although the findings of this study that is related to the access and usability of new technologies in patients diagnosed with psychotic disorders are very similar to the data obtained in the previous studies conducted in patients with psychosis [14,17,18,24,31] and in studies conducted in the Spanish general population [23,33], larger studies are needed to generalize our results, based on a small sample, to a broader patient population with psychosis in Spain to confirm that they are a good target for eHealth interventions.

Second, our results showed that there is a widespread use of internet to obtain information about health, not just by patients diagnosed with psychotic disorders but also by the general population [23]. However, we would like to highlight the substantial negative experiences related to internet usage that we found in our sample. Due to the great extent of internet usage in our society, we believe that further studies focusing on how internet usage affects patients are needed to understand the effect that this resource has on these patients and to study its role as a risk factor for psychosis.

Finally, we did not find any differences between the patient groups in terms of interest in eHealth services, allowing us to conclude that regardless of the demographic or clinical characteristics of patients, they would be equally interested in these interventions. However, in every category measured, we found 20%-30% of patients who were systematically "not interested" in the interventions suggested. As it has been shown in previous studies, personality can affect internet and mobile phone use [39,40]. In accordance with this, it would be interesting to replicate this study with a larger sample and to include specific measures of personality, interest, and patients' expectations because we believe that it would not be possible to achieve any promising results with the use of technology advances if the patients do not feel encouraged and motivated to use these resources. This is the reason why future investigations must focus on better understanding the patients' point of view to truly achieve a personalized measure of the patients' health status.

This study is the first approach to such patients' perspective. We aimed to describe patients' current situation in terms of the availability of technology and the experiences and opinions related to its usage. However, further studies are needed.

Acknowledgments

This study was supported by Generalitat Valenciana (PROMETEO/2016/082 grant) and CIBERSAM and Fondo de Investigación Sanitaria (PI01399, PI13/00447; PI17/00402, Cofunded FEDER).

Authors' Contributions

LB, BL, MHV, CC, and ME recruited and evaluated the patients. LB, IB, DA, and AMGP drafted and discussed the paper. JS designed and supervised the project. LB and JS analyzed the data and the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instrument.
[PDF File (Adobe PDF File), 72 KB - mental_v5i3e51_app1.pdf]

Multimedia Appendix 2
Ethics and confidentiality. Information sheet for the patient.
[PDF File (Adobe PDF File), 34 KB - mental_v5i3e51_app2.pdf]

References

1. Perkins DO, Gu H, Boteva K, Lieberman JA. Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. Am J Psychiatry 2005 Oct;162(10):1785-1804. [doi: 10.1176/appi.ajp.162.10.1785] [Medline: 16199825]
2. Marshall M, Rathbone J. Early intervention for psychosis. Cochrane Database Syst Rev 2011 Jun 15(6):CD004718 [FREE Full text] [doi: 10.1002/14651858.CD004718.pub3] [Medline: 21678345]
3. Amos A. Assessing the cost of early intervention in psychosis: a systematic review. Aust N Z J Psychiatry 2012 Aug;46(8):719-734. [doi: 10.1177/0004867412450470] [Medline: 22696550]
4. Bonet L, Izquierdo C, Escartí MJ, Sancho JV, Arce D, Blanquer I, et al. Use of mobile technologies in patients with psychosis: A systematic review. Rev Psiquiatr Salud Ment 2017;10(3):168-178 [FREE Full text] [doi: 10.1016/j.rpsm.2017.01.003] [Medline: 28258835]
5. World Health Organization. 2007. Global Observatory for eHealth URL: http://www.who.int/goe/en/ [WebCite Cache ID 6u6WeR3D0]
6. Bucci S, Barrowclough C, Ainsworth J, Morris R, Berry K, Machin M, et al. Using mobile technology to deliver a cognitive behaviour therapy-informed intervention in early psychosis (Actissist): study protocol for a randomised controlled trial. Trials 2015;16(1):404 [FREE Full text] [doi: 10.1186/s13063-015-0943-3] [Medline: 26357943]
7. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: A Cognitive Neuroscience-Informed Mobile App Intervention to Enhance Motivated Behavior and Improve Quality of Life in Recent Onset Schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77 [FREE Full text] [doi: 10.2196/resprot.5450] [Medline: 27125771]
8. Garety PA, Ward T, Freeman D, Fowler D, Emsley R, Dunn G, et al. SlowMo, a digital therapy targeting reasoning in paranoia, versus treatment as usual in the treatment of people who fear harm from others: study protocol for a randomised controlled trial. Trials 2017 Nov 02;18(1):510 [FREE Full text] [doi: 10.1186/s13063-017-2242-7] [Medline: 29096681]
9. Sanjuán J. ¿Tratar la mente o tratar el cerebro?. 1st edition. Bilbao, Spain: Desclèe De Brouwer; 2016.
10. Depp CA, Harmell AL, Vahia IV, Mausbach BT. Neurocognitive and functional correlates of mobile phone use in middle-aged and older patients with schizophrenia. Aging Ment Health 2016;20(1):29-35 [FREE Full text] [doi: 10.1080/13607863.2015.1008987] [Medline: 25768842]
11. National Alliance on Mental Illness (NAMI). Health and Technology Study 2014 URL: https://www.nami.org/About-NAMI/Publications-Reports/Survey-Reports/Health-and-Technology-Study-(2014) [accessed 2017-10-10] [WebCite Cache ID 6u6XgrKXh]
12. Smith A. The Smartphone difference: U.S. Smartphone Use in. Washington, DC: Pew Research Centre; 2015. URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015 [WebCite Cache ID 6u6XnvYrZ]
13. Tobitt S, Percival R. Switched on or switched off? A survey of mobile, computer and Internet use in a community mental health rehabilitation sample. J Ment Health 2017 Jul 04:1-7. [doi: 10.1080/09638237.2017.1340623] [Medline: 28675329]
14. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile Phone Ownership and Endorsement of "mHealth" Among People With Psychosis: A Meta-analysis of Cross-sectional Studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
15. Thomas N, Foley F, Lindblom K, Lee S. Are people with severe mental illness ready for online interventions? Access and use of the Internet in Australian mental health service users. Australas Psychiatry 2017 Jun;25(3):257-261. [doi: 10.1177/1039856217689913] [Medline: 28139947]
16. Gitlow L, Abdelaal F, Etienne A, Hensley J, Krukowski E, Toner M. Exploring the Current Usage and Preferences for Everyday Technology among People with Serious Mental Illnesses. Occupational Therapy in Mental Health 2017 Mar;33(1):1-14 [FREE Full text] [doi: 10.1080/0164212X.2016.1211061]
17. Trefflich F, Kalckreuth S, Mergl R, Rummel-Kluge C. Psychiatric patients' internet use corresponds to the internet use of the general public. Psychiatry Res 2015 Mar 30;226(1):136-141. [doi: 10.1016/j.psychres.2014.12.037] [Medline: 25623020]
18. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital Technology Use Among Individuals with Schizophrenia: Results of an Online Survey. JMIR Ment Health 2016;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]

19. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of Interventions Delivered Online and Through Mobile Phones for People Who Experience Severe Mental Health Problems: A Systematic Review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
20. Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Dec;11(4):290-295. [doi: 10.1111/eip.12237] [Medline: 25808317]
21. Ben-Zeev D, Scherer EA, Gottlieb JD, Rotondi AJ, Brunette MF, Achtyes ED, et al. mHealth for Schizophrenia: Patient Engagement With a Mobile Phone Intervention Following Hospital Discharge. JMIR Ment Health 2016;3(3):e34 [FREE Full text] [doi: 10.2196/mental.6348] [Medline: 27465803]
22. Kannisto KA, Adams CE, Koivunen M, Katajisto J, Välimäki M. Feedback on SMS reminders to encourage adherence among patients taking antipsychotic medication: a cross-sectional survey nested within a randomised trial. BMJ Open 2015 Nov 09;5(11):e008574 [FREE Full text] [doi: 10.1136/bmjopen-2015-008574] [Medline: 26553830]
23. Instituto Nacional de Estadística. Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares. 2016 URL: http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176741&menu=ultiDatos&idp=1254735976608 [WebCite Cache ID 6u6YxYOTE]
24. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do We Still Have a Digital Divide in Mental Health? A Five-Year Survey Follow-up. J Med Internet Res 2016 Nov 22;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
25. Miller BJ, Stewart A, Schrimsher J, Peeples D, Buckley PF. How connected are people with schizophrenia? Cell phone, computer, email, and social media use. Psychiatry Res 2015 Feb 28;225(3):458-463. [doi: 10.1016/j.psychres.2014.11.067] [Medline: 25563669]
26. Borzekowski DL, Leith J, Medoff DR, Potts W, Dixon LB, Balis T, et al. Use of the internet and other media for health information among clinic outpatients with serious mental illness. Psychiatr Serv 2009 Sep;60(9):1265-1268. [doi: 10.1176/ps.2009.60.9.1265] [Medline: 19723745]
27. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, ICD-10. United States: World Health Organization; 2012.
28. Aref-Adib G, O'Hanlon P, Fullarton K, Morant N, Sommerlad A, Johnson S, et al. A qualitative study of online mental health information seeking behaviour by those with psychosis. BMC Psychiatry 2016 Dec 11;16:232 [FREE Full text] [doi: 10.1186/s12888-016-0952-0] [Medline: 27400874]
29. Baup H, Verdoux H. Frequency and pattern of Internet use in patients with schizophrenia or bipolar disorders seeking medical information. Psychiatry Res 2017 Jan;247:152-154. [doi: 10.1016/j.psychres.2016.11.028] [Medline: 27893996]
30. Digital Marketing Trends (DiTrendia) M. Informe mobile en España y en el mundo. 2016. 2016 URL: http://www.amic.media/media/files/file_352_1050.pdf [accessed 2018-07-12] [WebCite Cache ID 70qxZdgEr]
31. Lal S, Dell'Elce J, Tucci N, Fuhrer R, Tamblyn R, Malla A. Preferences of Young Adults With First-Episode Psychosis for Receiving Specialized Mental Health Services Using Technology: A Survey Study. JMIR Ment Health 2015;2(2):e18 [FREE Full text] [doi: 10.2196/mental.4400] [Medline: 26543922]
32. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]
33. Asociación PLIDMDC(. Encuesta AIMC a usuarios de internet. 2017 URL: http://www.aimc.es/-Navegantes-en-la-Red-.html [accessed 2017-10-19] [WebCite Cache ID 6uKQY0czm]
34. Nowland R, Necka EA, Cacioppo JT. Loneliness and Social Internet Use: Pathways to Reconnection in a Digital World? Perspect Psychol Sci 2018 Jan;13(1):70-87. [doi: 10.1177/1745691617713052] [Medline: 28937910]
35. Heinz A, Deserno L, Reininghaus U. Urbanicity, social adversity and psychosis. World Psychiatry 2013 Oct;12(3):187-197 [FREE Full text] [doi: 10.1002/wps.20056] [Medline: 24096775]
36. Alvarez-Jimenez M, Priede A, Hetrick SE, Bendall S, Killackey E, Parker AG, et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012 Aug;139(1-3):116-128. [doi: 10.1016/j.schres.2012.05.007] [Medline: 22658527]
37. Berry N, Bucci S, Lobban F. Use of the Internet and Mobile Phones for Self-Management of Severe Mental Health Problems: Qualitative Study of Staff Views. JMIR Ment Health 2017 Nov 01;4(4):e52 [FREE Full text] [doi: 10.2196/mental.8311] [Medline: 29092809]
38. Bhugra D, Tasman A, Pathare S, Priebe S, Smith S, Torous J, et al. The WPA-Lancet Psychiatry Commission on the Future of Psychiatry. Lancet Psychiatry 2017 Oct;4(10):775-818. [doi: 10.1016/S2215-0366(17)30333-4] [Medline: 28946952]
39. Lachmann B, Sindermann C, Sariyska RY, Luo R, Melchers MC, Becker B, et al. The Role of Empathy and Life Satisfaction in Internet and Smartphone Use Disorder. Front Psychol 2018;9:398 [FREE Full text] [doi: 10.3389/fpsyg.2018.00398] [Medline: 29636714]
40. Stodt B, Brand M, Sindermann C, Wegmann E, Li M, Zhou M, et al. Investigating the Effect of Personality, Internet Literacy, and Use Expectancies in Internet-Use Disorder: A Comparative Study between China and Germany. Int J Environ Res Public Health 2018 Mar 23;15(4) [FREE Full text] [doi: 10.3390/ijerph15040579] [Medline: 29570663]

Abbreviations

CP: chronic psychosis
EI: early intervention
EP: early psychosis
FEP: first episode of psychosis
NAMI: National Alliance on Mental Illness

Edited by J Torous; submitted 26.01.18; peer-reviewed by N Berry, K Donoghue; comments to author 03.03.18; revised version received 26.04.18; accepted 30.05.18; published 25.07.18.

Please cite as:
Bonet L, Llácer B, Hernandez-Viadel M, Arce D, Blanquer I, Cañete C, Escartí M, González-Pinto AM, Sanjuán J
Differences in the Use and Opinions About New eHealth Technologies Among Patients With Psychosis: Structured Questionnaire
JMIR Ment Health 2018;5(3):e51
URL: http://mental.jmir.org/2018/3/e51/
doi:10.2196/mental.9950
PMID:30045835

Original Paper

Using Smartphone Apps to Promote Psychiatric Rehabilitation in a Peer-Led Community Support Program: Pilot Study

Nora E Mueller[1], BA; Trishan Panch[2], MD, MPH; Cathaleene Macias[3], PhD; Bruce M Cohen[1,3], MD, PhD; Dost Ongur[1,3], MD, PhD; Justin T Baker[1,3], MD, PhD

[1]Psychotic Disorders Division, McLean Hospital, Belmont, MA, United States

[2]Wellframe Inc, Boston, MA, United States

[3]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Justin T Baker, MD, PhD
Psychotic Disorders Division
McLean Hospital
115 Mill Street
Belmont, MA, 02478
United States
Phone: 1 6178553913
Email: jtbaker@partners.org

Abstract

Background: Management of severe and persistent mental illness is a complex, resource-intensive challenge for individuals, their families, treaters, and the health care system at large. Community-based rehabilitation, in which peer specialists provide support for individuals managing their own condition, has demonstrated effectiveness but has only been implemented in specialty centers. It remains unclear how the peer-based community rehabilitation model could be expanded, given that it requires significant resources to both establish and maintain.

Objective: Here, we describe the results from a study of one such program implemented within Waverley Place, a community support program at McLean Hospital, emphasizing psychiatric rehabilitation for individuals with severe and persistent mental illness, as well as describing the challenges encountered during the implementation of the program. Key questions were whether the patients could, and would, successfully use the app.

Methods: The smartphone app offered multiple features relevant to psychiatric rehabilitation, including daily task lists, activity tracking, and text messaging with peer specialists. A 90-day program of activities, goals, and content specific to the community support program was created on the basis of a prior pilot, in collaboration between members of the app development team (WellFrame), and peers, clinical, and research staff associated with the program. Hospital research staff recruited patients into the study, monitored peer and patient engagement, and handled all raw data acquired from the study.

Results: Of 100 people approached for the study, a total of 13 provided consent, of which 10 downloaded and used the app. Two patients were unable to complete the app installation. Five used the app regularly as part of their daily lives for at least 20 days of the 90-day program. We were unable to identify any specific factors (eg, clinical or demographic) that affected willingness to consent or engage with the app platform in the very limited sample, although the individuals with significant app use were generally satisfied with the experience.

Conclusions: Smartphone apps may become a useful tool for psychiatric rehabilitation, addressing both psychiatric and co-occurring medical problems. Individualizing functions to each patient and facilitating connection with a certified peer specialist may be an important feature of useful apps. Unlike prior reports emphasizing that patients with schizophrenia will adopt smartphone platforms, we found that implementation of digital tools into existing community support programs for severe and persistent mental illness has many challenges yet to be fully overcome to realize the potential benefits such apps could have to promote systematization and cost reduction for psychiatric rehabilitation.

(JMIR Ment Health 2018;5(3):e10092) doi:10.2196/10092

KEYWORDS

psychosis; smartphone; app

Introduction

The imperative for health care providers to engage patients as proactive partners in their own health care has never been greater. A wide variety of community-based programs for people with mental illness are now recognized as evidence-based practices, but their provision of care is usually limited to the program facility because it is cost-prohibitive for program staff to meet with patients in their homes, neighborhood locations, or work settings [1-4]. This limitation in the loci of service provision is problematic because, while the psychiatric symptoms of many people can be symptomatically controlled by medication and psychotherapy, they are often not able to recover to premorbid levels of functioning without additional support in the community [2]. Therefore, there is a need for a cost-effective, low-effort measure to bridge this gap in service.

Mobile technologies, such as smartphone apps, show promise for remotely guiding and supporting patients with mental illness on an ongoing basis in a convenient and cost-effective manner. Smartphones, mobile phones that are able to access the internet and use apps, have become ubiquitous in the United States, and statistics indicate that this holds true among people with severe mental illness [5-8]. Additionally, previous studies have determined that mobile apps are acceptable to end users [9]. A systematic review by Firth et al found that in studies of smartphone apps in patients with schizophrenia, there was overall high retention and use by participants, with an overall retention rate of 92% [10]. The review also found relatively high use of the smartphone apps by participants, with participants using the apps more than 85% of the time [10]. In a previous study by Ben-Zeev et al, a smartphone app designed to provide illness management support to people with schizophrenia was used effectively by patients [11]. Participants were able to log on and access the resources on the app, participants accessed the app on average five times a day and 62% used it to reach out to their treaters [11]. In some studies, clinician engagement rates have been shown to be lower than patient rates [12]. One app designed for middle-aged people with serious mental illnesses, which led them through a ten-session psychosocial intervention on their smartphone, found good acceptability of smartphone app use in this population [13]. Additionally, the absence of stigma around use of smartphone apps may be one advantage of this approach. Previous research has determined that smartphones are viewed as a nonstigmatizing way to deliver care as they are commonplace in mainstream society, as is the use of mobile apps [14].

Despite progress, it is not clear whether older adults diagnosed with a major mental illness, many of whom also have a chronic physical health condition, are capable of using a health-promotion smartphone app on a regular basis in the community and if they are willing to do so. The present study examined the acceptability, viability, and usage patterns of a smartphone app developed by a software company ("WellFrame"), with which one of the authors is affiliated (TP). The app was tailored to the clinical programming of a community support program affiliated with McLean Hospital. A previous study of found that the smartphone app was acceptable to the program's clients during a 30-day trial [15]. During that trial, all participants were successfully able to use the app, used the app an average of 94% of days in the study, and said they would continue to use the app if given the opportunity. In the present study, the trial period was extended to 90 days in order to determine whether individuals with severe and persistent mental illness would continue to sustain use of the app over a longer trial and whether acceptability and usage patterns would change, particularly given that usage of most apps drops off precipitously over time [15].

Methods

Study Sample

The site for this study was Waverley Place, an outpatient community support program for adults with mental illness affiliated with McLean Hospital. The majority of program members were aged 50 years or older, and most were diagnosed with a schizophrenia spectrum disorder or major affective disorder as a young adult. Daily activities offered by the program include mental health and addiction support groups co-led by staff and members, art and craft sessions, cooking classes, recreational and social activities, physical health promotion (exercise classes, trips to a local gym), counseling, and direct help with daily living tasks, as well as referrals to volunteer or paid work outside the program. Additionally, every member chooses a "contact person," a certified peer specialist, with whom they check in on a semiregular basis. The program employs both Master's level mental health staff and certified peer specialists. The program administrator is a psychiatrist at McLean Hospital.

Recruitment and Screening

This protocol was reviewed and approved by the Partners Healthcare Institutional Review Board and all participants gave informed consent. A convenience sample was recruited during usual group activities at the program, for example, support groups, program planning meetings, exercise sessions, social events, and impromptu gatherings for casual conversation and recreation. Typically, a study representative provided a brief description of the smartphone app ("*WellWave* ") and the app evaluation study, inviting those interested in learning more to meet with the study representative individually at a later time, where they demonstrated the app and explained study aims and procedures. Additionally, the study staff attended all community meetings to make announcements about the project and posted flyers with study staff contact information in order to stimulate recruitment. A total of 100 individuals were approached by research staff, over an 8-month period of enrollment. To ensure inclusiveness, the only eligibility criteria for study enrollment were (1) access to a mobile phone compatible with the app or willingness to use a study-provided one and (2) current membership in Waverley Place. Interested individuals who did not own an app-compatible phone were told they could receive a loaner phone as soon as these were available, and that the project would provide instruction in how to use the phone. Only 1 participant accepted a loaner phone.

Smartphone App Intervention

This pilot study focused on the two main components of the *WellWave* app: (1) to improve patient functioning and (2) to allow program staff to confidentially monitor the well-being of patients when they are not attending the program. The purpose of the app is to improve engagement of individuals in the rehabilitation program through increasing communication with their care provider and assigned peer specialist. A detailed description of the development and initial pilot of the app are reported by Macias et al [15]. Members of the Wellframe staff were actively involved in the planning and support phases of this study and provided the app free of charge for the duration of the pilot, as well as providing the phones for participants without smartphones to use. Members of the WellFrame staff did not, however, engage in any direct communication with the study participants, either in person or online, at any time during the study procedures.

Daily Tasks Lists

The *WellWave* app generates a list of suggested activities for each day in the 90-day study period using a predefined sequence that was fixed for all participants, such that a new "Daily Tasks" list would appear on the participant's phone screen every morning. These optional activities included internet links to recommended readings and videos; questionnaires including brief free-text reports on well-being (eg, "I am feeling...") or mood and symptom self-ratings; and exercise prompts, which consisted of a goal for daily walks (ie, steps or time-based) that gradually increased from 5 minutes to 20 minutes over the course of the 90-day program. Medication reminders also appeared on the Daily Tasks list for participants who choose to set up this option. An exercise reminder appeared on the screen every day. Testimonial or motivational videos created by the community support program's current peer staff appeared on the Daily Task list early in the 90-day program to encourage engagement.

The app automatically recorded each participant's selection of a task, how long the selected task remained open, and any response the participant made to the task prompt. These recordings were subsequently used to graph fluctuation in task engagement over time. For instance, the app electronically tracked which internet resources (eg, psychoeducational videos and readings) were activated, as well as the duration of engagement.

In addition, participants were asked to report (through the app) their emotional and physical well-being; relative ease of use, which could be visualized in a Web-based dashboard visible to the research and program staff; completion rates; and sensitivity to detect change over time in relation to life events reported by each user. These self-report measures included standardized self-assessment instruments, including brief versions of the PROMIS scales [16], single-item global questions commonly used in medical settings (eg, "How would you describe your physical health?" where 1=poor to 5=excellent), and brief mood and symptom checklists. A total of 2 to 6 tasks appeared on the app's Daily Tasks screen on any given day.

Text Messages

The *WellWave* app offered two-way text messaging between app users and community program peer staff (certified peer counselors diagnosed with a mental illness). This messaging was confidential and HIPAA-compliant because all text conversations took place exclusively within the app. The app preserves this text, ordered by sender id, date, and time of day. In this pilot study, we were primarily interested in measuring how often such electronic contact took place, and how often the messaging was reciprocal (ie, a two-way conversation between peer staff and app user).

Measures

Intake and Exit Interviews

Questionnaires were administered at the time of enrollment and in a poststudy exit interview. Intake questions were designed to measure (1) prior experience with smartphones; and (2) current mental and physical health problems that might interfere with app use. The latter measure of self-perceptions used the same wording and scaling as measures in the *WellWave* app. Questionnaires were offered as a choice between paper or computer formats, which used REDcap. The exit interview questionnaire also contained several rating scale measures of satisfaction with the app, and inquired about personal reasons for using or not using the app. The app feedback questionnaire was posed as a series of prompts, such as "I still find it hard to use WellWave without asking for help" and "I'd recommend WellWave to a friend" rated on a 1 to 5 Likert scale. A member of the research team also conducted interviews with each of the participants during the exit interview. This interview consisted of 12 questions about specific features the participants liked and didn't like about the app, and whether or not they would continue the app if given the option.

Engagement With the Smartphone App

Ease of use and acceptability were measured quantitatively, including items for whether app training was successfully completed (yes or no), how frequently the app was used, how intensively each of the app components was used, and duration of app use during a common 90-day study period.

Frequency of app use was operationally defined as a count of study days the app was used (turned on) during each participant's 90-day study period.

Intensity of app use was defined as the ratio of total tasks opened to total tasks viewed in the Daily Tasks list across all days the app was used. *Intensity* was calculated for each participant as one summary percentage, and as separate percentages for specific types of tasks. Because only one participant chose to use the medication reminder option, summary measures of use *frequency* and *intensity* were based only on prompts in the Daily Tasks lists that reminded participants to exercise, offered links to videos and readings, or asked for reports of well-being.

Duration of app use was defined as a count of calendar days between first and last day the app was used by each study participant.

Results

Participants

Figure 1 outlines the participant flow through the study. Of the 100 people eligible and approached to participate in the study, a total of 13 were recruited. Members who met study eligibility criteria entered the project gradually over an 8-month open enrollment period (August 31, 2015 to March 4, 2016). However, only ten participants actually downloaded the app and were onboarded to the study. Of the three that never downloaded the app, 2 were excluded due to incompatibility of their personal smartphone; and 1 was lost to follow up after consent. Of the remaining 10 participants two individuals were unable to use the app after intake training. Three successfully completed app training, but then used the app only briefly. Five participants used the app on 20 or more study days (range 20-92 days). All 13 participants completed baseline assessments and were therefore included in analyses of baseline data.

All the participants were non-Hispanic Caucasian. A majority were over the age of 40 years (n=8), and about half were over the age of 50 years (n=6). Most of the participants were women (n=10). The women were more diverse in age (range 29-63 years), and the men were aged 52-60 years. The sample was split between those diagnosed with a schizophrenia spectrum disorder (n=5), and those with a major affective disorder (n=8), either bipolar disorder or severe depression. Most of the sample (n=7) also reported being under treatment for a chronic physical condition (eg, heart disease, asthma, arthritis, cancer). Psychiatric and physical health diagnoses were evenly distributed across age groups and sexes. As is true of the program membership in general, the study sample was well-educated: all except 1 study enrollee had some postsecondary education, and 9 of the 10 had completed 2 or more years of college. Two of the 10 were already in a previously published pilot [14].

The following case descriptions of study participants include identified reasons for not using the app, and explicit or inferred attitudes expressed by participants toward the app. These qualitative explanations for app use or nonuse were derived from researcher-recorded observations during and after app training, participant verbal and written feedback in exit interviews, text messages sent to staff regarding app use, and participants' free-text responses to app-generated questions about their health and daily activities. Additionally, participants total messages received and sent are displayed in Table 1.

Unable to Use App

Two participants (both women) who were unable to use the app at all after training. One (aged 61 years, bipolar disorder) had difficulty staying awake during app training, was hospitalized for psychiatric reasons a few days later, and remained hospitalized for the duration of the study. The other (aged 63 years, schizophrenia) accepted a loaner phone in order to be eligible for the project, but she had minimal experience using a smartphone, and after 3 instructional sessions, gave up trying to learn to use the app. Both participants had earned college degrees (Master's degree and registered nurse) as young adults, were long-term members of Waverley Place, and checked psychotic symptoms and cognitive problems on intake questionnaires.

App Training Only

One man (aged 52 years, major depression) and two women (aged 24 years, bipolar disorder and aged 59 years, schizophrenia) completed intake training but only used the app for at most five days. All 3 were relatively new to the community support program. In their exit interviews, the man (age 52 years) said, "I forgot I had the app." The older woman (age 59 years) said she had accidentally deleted the app. The 24-year old woman was lost to follow-up soon after enrollment.

App Engaged

Of the 13 participants recruited, 5 study participants used the app for 20 or more of their 90 study days. Observations made by study staff and program staff about individual participants utilization of the app are described in Table 2.

Associations Between Participant Characteristics and Level of App Engagement

The sample was homogeneous with regard to race (all non-Hispanic Caucasian), education (12-16 school years), prior experience using a mobile phone (all except one had used a smartphone for 2-4 years), and current level of exercise (all had some scheduled physical activity). Due to the small sample size we could not test whether any of these four background characteristics had any impact on app use. However, from a general examination of the data there does not seem to be any associations.

Participant Feedback

Five participants completed an exit interview at the end of the 90-day study period. In contrast to an earlier 30-day pilot study, where all participants reported satisfaction with the app [14], in the present study only 2 participants said they would continue to use the app if given the chance to do so. Although, one participant who reported that they would not want to continue using the app, cited frustration with the technical malfunctions of the exercise portion of the app, where the exercise tracker would stop working mid-walk, as their main source of dissatisfaction and stated that "...without the exercise part, the videos were good, and the med reminders were good." Additionally, 4 of the 5 participants indicated that they enjoyed receiving messages from the peer specialists and that these messages made them feel more connected to the staff. Participant's suggestions for improving the app included adding a social feature to interact with other members.

Figure 1. Outline of participant flow through the study.

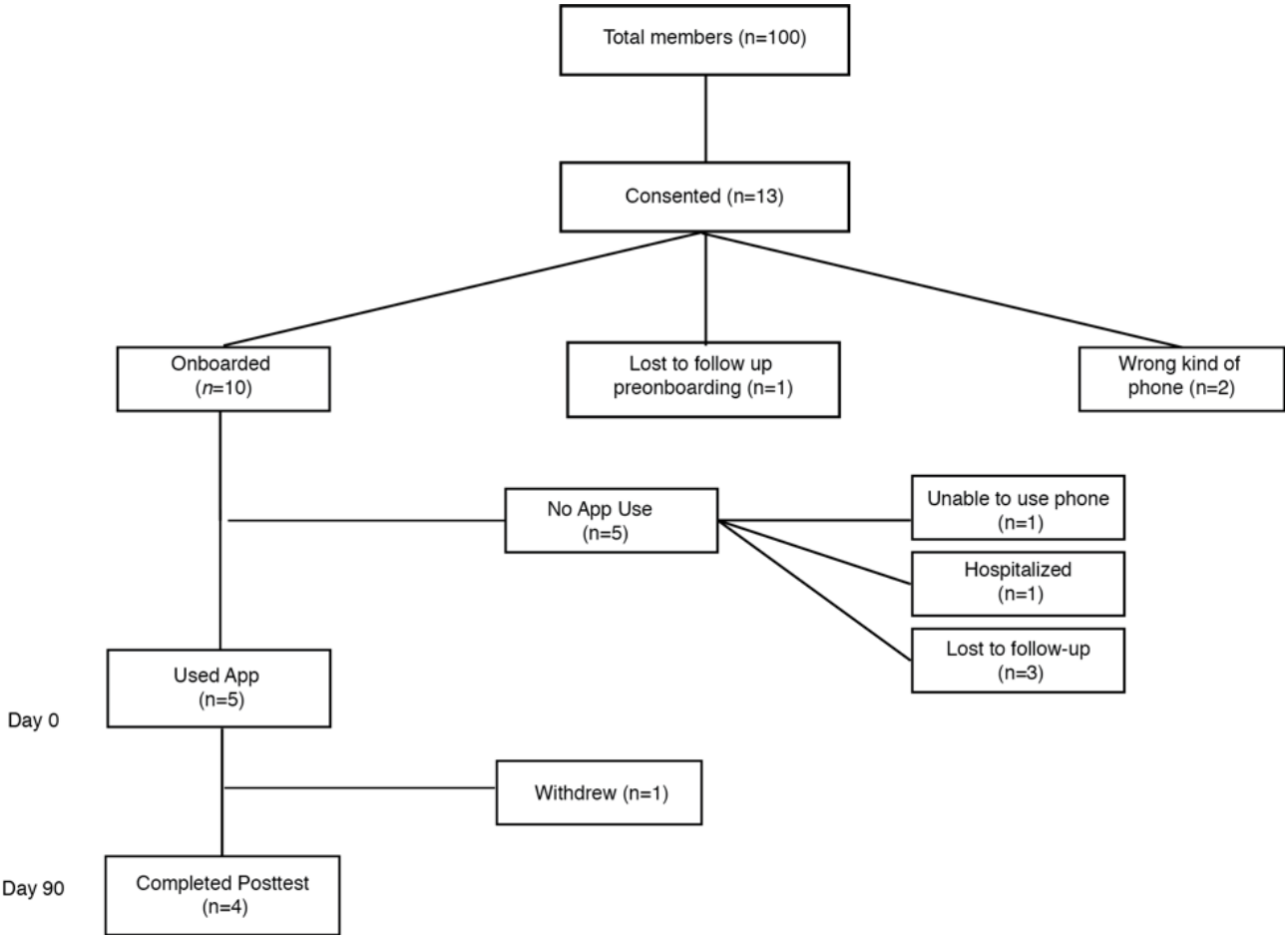


Table 1. Frequency of messaging of participants and staff members.

Subject ID	Number of user messages	Number of staff messages
WW013	0	10
WW012	1	7
WW010	0	11
WW009	5	23
WW007	0	14
WW006	2	23
WW005	14	34
WW004	0	0
WW002	4	16
WW001	53	51

Table 2. Case descriptions for the 5 participants who actively used the app beyond onboarding.

Characteristics (age, gender)	Diagnosis	Key app use features	App use difficulties
Late 40s, female	Schizophrenia	While active, they engaged in almost all the app digital components (eg, surveys, videos, readings) except medication management, and took 20 app-recorded walks during her 28 days of app use.	Hand tremors made it difficult to key responses into their phone, but they did not mention this as an obstacle to app use. The reason they gave for withdrawing from the project was that the app surveys designed to track her moods and psychiatric symptoms "made me feel bad" and "more symptomatic."
Late 20s, female	Schizoaffective Disorder	This participant used the app about half the days she was enrolled in the project, and her days of use were sporadic across the 90-day study period.	They refused to complete both intake and exit questionnaires. The participant took only 3 recorded walks and expressed concern at intake that her exercise performance would be compared to that of other participants.
Early 40s, female	Bipolar Disorder	This participant remained steadily engaged in the app for the duration of the study in spite of several chronic health conditions (cardiovascular disease, diabetes, asthma), recurrent migraine headaches, and an acute episode of severe bronchitis. Their primary psychiatric symptoms were anxiety and depression, and she rated herself at intake as very lonely, suspicious, edgy, and easily upset. The participant reported that insomnia often left them tired and unable to concentrate, but they used the app almost daily, and responded to about half of the tasks they saw listed on the Daily Tasks screen. They also had the second-highest rate of conversational text-messaging with peer staff, and the second-highest count of app-recorded exercise (31 walks).	Not applicable.
Early 60s, male	Schizophrenia	This participant opened the app nearly every day to see what was listed in the Daily Tasks screen. They were also intensively engaged in all the app tasks except exercise (8 recorded walks). The participant responded to 56% of the tasks he viewed on the Daily Tasks list, completing about 70% of the app-generated surveys, and connecting to 70% of the internet readings and videos. The participant sent a text to peer staff saying "This is a good app!"	They exhibited hand tremors during app training and reported that sleep apnea reduced his ability to concentrate.
Early 60s, female	Bipolar Disorder	By the end of the project, this participant had viewed 90% of the internet readings and videos that appeared on the Daily Tasks lists, completed 84% of the self-report surveys, and they were the only participant to use the medication reminder option. In spite of needing a walker to get around, they recorded 58 walks on an elliptical machine in response to exercise prompts. This participant also responded to almost all the self-report surveys received on their phone and sent periodic free-text descriptions of events in their social life.	This participant was slow in keying in responses because of hand tremors.

Discussion

Principal Results

The aim of this study was to determine if extending the trial of a previously piloted smartphone app from 30 days to 90 days would affect the acceptability and use patterns. In the present study the uptake of the app was low: of the 100 members of the program approached to participate in the study, only 13 signed consent and only 10 downloaded the app onto their phone, which was not a problem encountered in the previous study. However, those who did use the app found it acceptable.

The current study is notable because of the recruitment difficulties encountered by the study staff. Despite the minimal inclusion criteria and efforts of the research staff to use multiple avenues to stimulate recruitment (one-on-one recruitment, flyers, group announcements) only 13 members of the 100 approached

ultimately consented to be in the study. Technology is often looked to as an exciting component for the future for physical and mental health care, but this study may be a demonstration that although conceptually mobile apps may provide a way to provide more or better services to people, and fulfill a need in health care, sometimes people do not want to use them. The present study was able to provide a better look at the desire for people to engage with these kinds of apps in a real-world situation than previous studies, because unlike in previous studies, participants were provided no monetary compensation for their participation [17]. Therefore, the participants had little other incentive to participate other than interest in the app itself.

There are certain characteristics of Waverley Place that may have contributed to the low adoption of the app in this population. Specifically, members of this day program already had a staff member who they were assigned to check in with them and provide support, reducing the need for an adjunctive digital support tool. Additionally, the day program is open 8 hours a day, 5 days a week and members are free to come in and receive support and guidance from a staff member at any time. Consequently, there may have been no need for this app to fill in this particular program. However, a similar app may do very well in a population where time and resources of the program and staff are more limited, and this app, or similar apps, may represent a way to maintain contact with members in ways they otherwise would not be able to. Therefore, as more smartphone apps are developed the target population should be carefully considered to determine whether a smartphone app would be wanted or useful.

People who are older may be less technologically literate, which was demonstrated in our sample by the difficulty study staff had teaching participants using study provided smartphones (several one-on-one sessions, before the participant ultimately withdrew). Participants in this study were older adults with over half being over the age 40, and this age range was representative of the program members, particularly given that those who attend the program on a regular basis are older, indicating that this may have limited the adoption and use of the app in our study. Additionally, many of the members of the program used government subsidized flip phones and did not want to use an additional device for the duration of the study, which is similar to results found in other studies in community support clinics [17]. Therefore, smartphone apps such as *WellWave* may be more successfully accepted in a population where people are already using a smartphone, such as in a population of younger adults, in an early intervention or first episode clinic. For example, in a study published by Kumar et al) a majority of their potential participant pool did not have a smartphone to begin with, but they were able to effectively able to distribute smartphones for use during their intervention, potentially due to using an Early Psychosis population [17]. Interestingly, Ben-Zeev et al was able to effectively train and deploy study supplied phones in their intervention, in a study with a mean age of 45.9 years, similar to that of the present study [11]. Therefore, there may be more factors impacting a person's interest in using an app other than age, and technological literacy, which is commonly associated with age, that need to be explored in order to develop a strategy for the most effective development and implementation of these resources.

Implications for the Design and Evaluation of Mobile Phone Apps

Smartphone apps that require interactive responses to multiple features appear to be appropriate for most adults with a major mental illness who have smartphone experience, 80% (8/10) of study enrollees who downloaded the app were able to learn to use the app, and 50% (5/10) found the app acceptable for regular use in their everyday lives. However, the personal reports suggest that use and comfort with specific features differs greatly among individual patients.

Participants who used the app regularly had moderate to high response rates for every app component, except the option to receive medication reminders. Most study enrollees had been coping with mental illness for many years, so it is likely they already had medication management systems in place and, unlike new patients, may not need these reminders. Alternatively, medication use is often a complex negotiation between patient and clinician, and an app may not be the best way to assist compliance for many users.

Smartphone apps, like *WellWave*, appear especially useful for community support programs dedicated to promoting autonomy and community living because patient well-being can be assessed frequently and efficiently, at low cost and with minimal staff time.

More research is needed to determine which app features are most important to adults with mental illness and their mental health service providers, as well as what level of app complexity is a good fit with particular groups of people who share the same psychiatric symptoms or educational attainment. For instance, the participant who withdrew after a month in the project said that she found the negatively-worded symptom-assessment questions particularly distressing. Ideally, smartphone apps should be pilot-tested continuously in small sample studies while they are still being developed so that app features can be added and tested in a stage-wise fashion, with one or two new components added at a time, beginning with a streamlined menu of app activities and progressing toward a more demanding design. This incremental approach would reveal at what level the app becomes unacceptable or unusable by people with differing levels of functioning and cognitive clarity and allow the eventual production of separate versions of the same app tailored to the needs of specific types of people. Beyond types and groups, symptoms, styles, and needs are very different from one person to the next, and app design and choice may need to be individualized.

Additional research should assess what factors impact the adoption of a smartphone app in a population. Apps and other technology are looked to as potential ways to disseminate quality mental health care more efficiently, but as the present study illustrates more work needs to be done to understand what interventions will work and will be accepted in a given population.

Limitations

This study was limited by a small sample size, which may limit the generalizability of our findings to other clinical samples and treatment settings. Nonetheless, our experiences provide some unexpected insight into the demand for smartphone apps in a population like Waverley Place when no other financial incentive is given. The small sample size did not allow us to draw any conclusions about the participant characteristics that most influenced app use. Although some anecdotal data was collected by study staff about the reasons members of the program did not want to participate, further research is warranted to explore more systematically the reasons and characteristics of people who do and do not want to participate in these app studies. This would have important implications for development and dissemination of future apps to support individuals with severe and persistent mental illness.

Conclusions

A smartphone app designed to promote and track emotional well-being can become a useful tool for many community-based patients served by mental health programs. All adults diagnosed with a major mental illness should be offered the opportunity to learn to use a health-promotion smartphone app even if they are psychiatrically symptomatic, past the age of 50 years old, or coping with a cognitive impairment or a disabling physical health condition. However, this study illustrates the importance of researchers being cognizant of factors that may affect the implementation of such apps in existing community support programs, such as how features in the app complement or overlap with existing services, and the extent to which the application is offered as a research opportunity versus an integral part of the program's support services.

Conflicts of Interest

None declared.

References

1. O'Hara K, Stefancic A, Cabassa LJ. Developing a peer-based healthy lifestyle program for people with serious mental illness in supportive housing. Transl Behav Med 2017 Dec;7(4):793-803 [FREE Full text] [doi: 10.1007/s13142-016-0457-x] [Medline: 28155109]
2. Bouvet C, Battin C, Le Roy-Hatala C. [The Clubhouse model for people with severe mental illnesses: Literature review and French experiment]. Encephale 2015 Dec;41(6):477-486. [doi: 10.1016/j.encep.2014.09.001] [Medline: 25438970]
3. Væggemose U, Lou S, Frumer M, Christiansen NLS, Aagaard J, Ørtenblad L. Community Families: Bridging the gap between mental health services and civil society - A qualitative study from users' perspective. Int J Soc Psychiatry 2017 Mar;63(2):132-138. [doi: 10.1177/0020764016687789] [Medline: 28110623]
4. Drake R, Becker D, Biesanz J, Wyzik P, Torrey W. Day treatment versus supported employment for persons with severe mental illness: a replication study. Psychiatr Serv 1996 Oct;47(10):1125-1127. [doi: 10.1176/ps.47.10.1125] [Medline: 8890346]
5. Proudfoot J. The future is in our hands: the role of mobile phones in the prevention and management of mental disorders. Aust N Z J Psychiatry 2013 Feb;47(2):111-113. [doi: 10.1177/00048674124714441] [Medline: 23382507]
6. International Telecommunications Union. 2011. The World in 2011: ICT Facts and Figures URL: http://www.itu.int/ITU-D/ict/facts/2011/index.html [WebCite Cache ID 718F8bW7h]
7. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
8. Ben-Zeev D, Kaiser S, Brenner C, Begale M, Duffecy J, Mohr D. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
9. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. Telemed J E Health 2015 Mar;21(3):213-216 [FREE Full text] [doi: 10.1089/tmj.2014.0105] [Medline: 25536190]
10. Firth J, Torous J. Smartphone Apps for Schizophrenia: A Systematic Review. JMIR Mhealth Uhealth 2015 Nov 06;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
11. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
12. Španiel F, Hrdlička J, Novák T, Kožený J, Höschl C, Mohr P, et al. Effectiveness of the information technology-aided program of relapse prevention in schizophrenia (ITAREPS): a randomized, controlled, double-blind study. J Psychiatr Pract 2012 Jul;18(4):269-280. [doi: 10.1097/01.pra.0000416017.45591.c1] [Medline: 22805901]
13. Whiteman K, Lohman M, Gill L, Bruce M, Bartels S. Adapting a Psychosocial Intervention for Smartphone Delivery to Middle-Aged and Older Adults with Serious Mental Illness. Am J Geriatr Psychiatry 2017 Aug;25(8):819-828. [doi: 10.1016/j.jagp.2016.12.007] [Medline: 28169129]

14. Palmier-Claus JE, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry 2013 Jan 23;13:34 [FREE Full text] [doi: 10.1186/1471-244X-13-34] [Medline: 23343329]
15. Macias C, Panch T, Hicks YM, Scolnick JS, Weene DL, Öngür D, et al. Using Smartphone Apps to Promote Psychiatric and Physical Well-Being. Psychiatr Q 2015 Dec;86(4):505-519. [doi: 10.1007/s11126-015-9337-7] [Medline: 25636496]
16. Christo G, Jones SL, Haylett S, Stephenson GM, Lefever RMH, Lefever R. The Shorter PROMIS Questionnaire: further validation of a tool for simultaneous assessment of multiple addictive behaviours. Addict Behav 2003 Mar;28(2):225-248. [Medline: 12573676]
17. Kumar D, Tully LM, Iosif A, Zakskorn LN, Nye KE, Zia A, et al. A Mobile Health Platform for Clinical Monitoring in Early Psychosis: Implementation in Community-Based Outpatient Early Psychosis Care. JMIR Ment Health 2018 Feb 27;5(1):e15 [FREE Full text] [doi: 10.2196/mental.8551] [Medline: 29487044]

Edited by J Torous; submitted 11.02.18; peer-reviewed by J Firth, O Byambasuren, D Fulford; comments to author 23.03.18; revised version received 18.05.18; accepted 29.05.18; published 15.08.18.

Please cite as:
Mueller NE, Panch T, Macias C, Cohen BM, Ongur D, Baker JT
Using Smartphone Apps to Promote Psychiatric Rehabilitation in a Peer-Led Community Support Program: Pilot Study
JMIR Ment Health 2018;5(3):e10092
URL: http://mental.jmir.org/2018/3/e10092/
doi:10.2196/10092
PMID:30111526

Review

Digital Technology for Caregivers of People With Psychosis: Systematic Review

Juliana Onwumere[1,2], PhD; Filipa Amaral[1], MSc; Lucia R Valmaggia[1], PhD

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

[2]Bethlem Royal Hospital, South London and Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Juliana Onwumere, PhD
Institute of Psychology, Psychiatry and Neuroscience
Department of Psychology
King's College, London
16 De Crespigny Park
Denmark Hill London,
United Kingdom
Phone: 44 0207 848 0197
Email: juliana.1.onwumere@kcl.ac.uk

Abstract

Background: Psychotic disorders are severe mental health conditions that adversely affect the quality of life and life expectancy. Schizophrenia, the most common and severe form of psychosis affects 21 million people globally. Informal caregivers (families) are known to play an important role in facilitating patient recovery outcomes, although their own health and well-being could be adversely affected by the illness. The application of novel digital interventions in mental health care for patient groups is rapidly expanding; interestingly, however, far less is known about their role with family caregivers.

Objective: This study aimed to systematically identify the application of digital interventions that focus on informal caregivers of people with psychosis and describe their outcomes.

Methods: We completed a search for relevant papers in four electronic databases (EMBASE, MEDLINE, PsycINFO, and Web of Science). The search also included the Cochrane database and manual search of reference lists of relevant papers. The search was undertaken in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses reporting guidelines.

Results: The search identified 9 studies derived from 8 unique datasets. Most studies were assessments of feasibility and were undertaken in the United States. Interventions were predominately Web-based, with a focus on improving the caregivers' knowledge and understanding about psychosis.

Conclusions: This study offers preliminary support for the feasibility and acceptability of digital interventions for psychosis in informal caregiver populations. However, the findings underpin a clear need for greater development in the range of caregiver-focused digital approaches on offer and robust evaluation of their outcomes. The use of digital approaches with caregiver populations seemingly lags someway behind the significant developments observed in patient groups.

(JMIR Ment Health 2018;5(3):e55) doi:10.2196/mental.9857

KEYWORDS

carers; digital interventions; families; psychosis; technology

Introduction

Informal Care Provision

Despite optimal pharmacological interventions, as many as one-third of patients with psychosis might continue to experience persistent psychotic symptoms, as well as comorbid conditions, including depression and anxiety [1,2], implying that across the globe, as care provision for adults living with psychosis conditions continues to devolve from hospitals to community settings, informal (unpaid) caregivers provide the bulk of care and play a central role in affecting treatment outcomes. Informal caregivers are primarily close relatives of patients, such as their parents, siblings, partners, and offspring, and many live with and maintain close regular contact with their relatives, particularly during the early illness course [3,4]. Patients who are supported by informal caregiving relationships

tend to have improved functioning and recovery outcomes, including improved life expectancy [5], fewer relapses and need for hospital admissions [6], and higher levels of engagement with and improvements from prescribed treatments [7,8]. The economic value of unpaid caregiver support is estimated at several billion pounds each year [9]; these figures are remarkable in the light of robust findings indicating that high levels of burden are reported by caregivers and at least one-third admit to being at "breaking point" in their role [10,11]. Approximately 40% of caregivers report clinical levels of depression and anxiety [4,12,13]; feelings of loss, grief, and despair are also commonplace [14]. The physical health of caregivers can also be compromised, including experiencing elevated rates of sleep disturbance [15-17].

Interventions to improve caregivers' understanding of psychosis, facilitate adaptive coping strategies, and provide support and stress management skills have proven efficacy [18,19] and are included in treatment guidelines in many regions, including Canada [20], Australia [21], the United States [22], and the United Kingdom [23]. However, there remains an ongoing issue of how to best increase the provision and access to evidence-based interventions for caregivers. Psychosis caregivers are a neglected group that has independent care, information, and support needs that mental health providers can typically struggle to respond. Reportedly, evidence-based family interventions, like other psychological therapies, are not widely available and rates of implementation in mental health trusts and services can range between 0% and 53% [24,25]. Several barriers to implementation and widening access have been posited, including issues related to family engagement, time demands, and insufficient staff and resources [26]. Hence, a growing need exists to explore options that help to address these obstacles and increase caregiver access to support interventions. Furthermore, the importance of seeking to identify effective and acceptable approaches to responding to the needs of psychosis caregivers is widely acknowledged, given the integral role they play in optimizing patient outcomes. Moreover, as caregivers with poor health status are more likely to relinquish their caregiving role and consequently affect patient outcomes and care costs, the need to focus on caregivers and optimal care provision is axiomatic.

Digital Interventions in Health Care

An ever-increasing proportion of the world's adult population is online [27]. In the United Kingdom, for example, 78% of the adult population (approximately 39.3 million) is online each day, with a similar proportion accessing the internet using mobile devices, including mobile phone or portable computers [28]. Searching for information, including those related to health issues, constitute some of the most popular Web-based activities [28,29]. The last decade has witnessed considerable growth and innovation in the application of digital technologies (electronic health), including virtual and augmented reality, to support the assessment, understanding, and treatment of a wide range of health conditions [30-33], including mental health, such as psychosis [34]. These have included, for example, developments with mobile apps (eg, mobile phones) to assess mood functioning and symptoms [35], Web-based psychological therapies [36], interactive short message service text messages [37], computerized interventions [38], and wearable technologies that offer real-time feedback on well-being and functioning such as activity and sleep quality [39,40]. Furthermore, digital interventions could be a useful way to offer time and cost-effective approaches to reach and engage with larger populations, including those who might be less willing or able to access standard services because of geography and travel burden, or where flexible modes of access and privacy are prioritized.

Study Aims

This study aimed to review the application of digital interventions and their outcomes with families (informal caregivers) of people with psychosis. It specifically aimed to characterize the type of interventions used with caregiver populations and their key components. This study included a broad definition of "digital interventions" to capture any approach designed to affect an individual's understanding, functioning, behavior, and well-being.

Methods

Design

This was a systematic review of the literature with a qualitative synthesis of the findings.

Selection Procedure

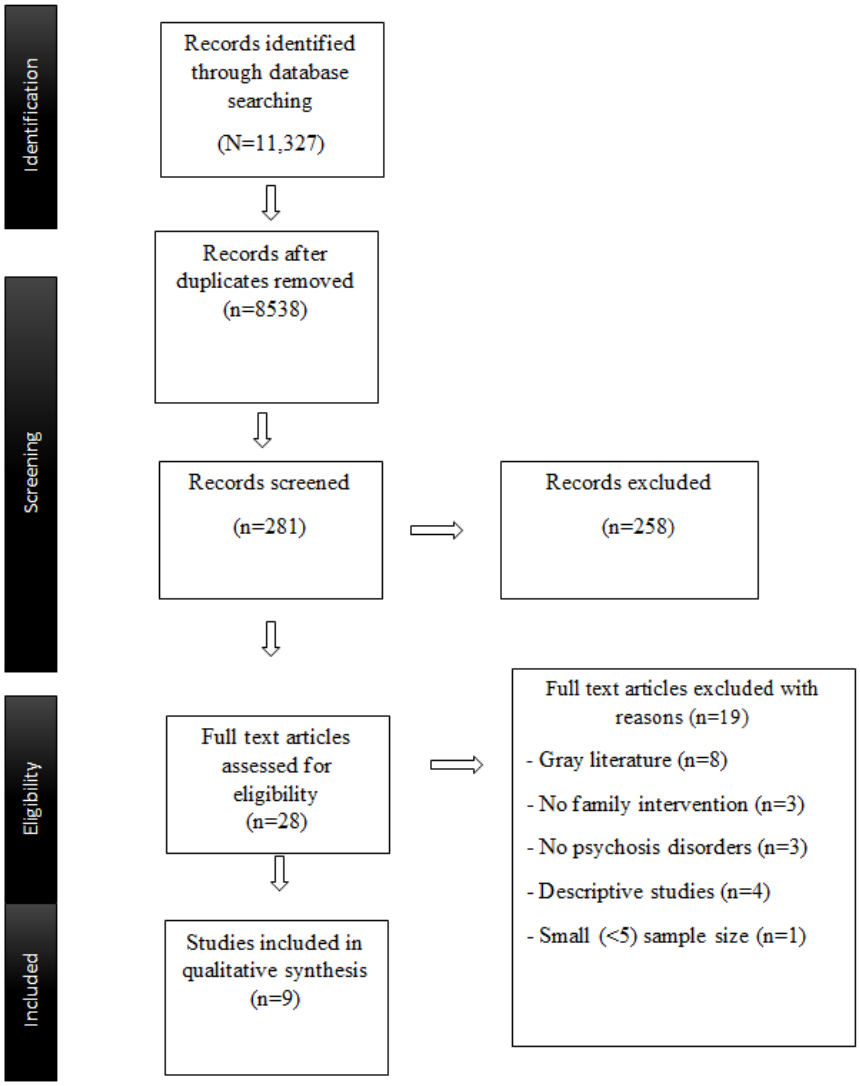
In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [41] we searched four electronic databases (EMBASE, MEDLINE, PsycINFO, and Web of Science) from inception to June 30, 2017. The search included the Cochrane database and manual search of the reference list of relevant papers.

Studies eligible for review were as follows: (1) those reporting the application and outcomes of a digital (ie, internet-, mobile-, virtual or augmented reality-, telephone-, or app-based) intervention; (2) those having target population including informal caregivers (ie, unpaid relatives or friends) of an individual with psychotic disorder; (3) those reporting caregiver-focused outcomes; (4) those published in English; (5) those reported in a peer-reviewed journal. Ineligible studies were those reporting data from a single person case studies and reviews, having samples with 5 or less participants, and including nonpsychotic disorder illness conditions. However, studies using mixed diagnostic groups were included if psychotic disorders constituted at least 50% of the sample. Furthermore, studies using patient and caregiver samples were eligible, although we focused on caregiver outcomes.

Search Criteria

To increase the search capabilities and accurate selection of papers, a comprehensive list of keywords and Medical Subject Headings, were used along with relevant search truncations and wildcards to capture variations in language and database indexing.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.



Search terms related first to the digital technologies comprising "virtual reality" OR teleme* OR telemedicine OR telepsychiatry OR telehealth OR eHealth OR mHealth OR "mobile phone" OR "mobile health" OR "mobile technolog*" OR "mobile application" OR smartphone* OR internet OR online OR "online system*" OR "social media" OR "Web-based intervention*" OR "augmented reality" OR e-learning OR computer* OR "computer assisted therapy" OR apps OR "mobile application." The second group of terms focused on caregivers comprising family OR families OR sibling*OR relative* OR "first-degree relative" OR partner*OR "domestic partner*" OR parent* OR caregiver* OR carer*. The third term related to the psychosis spectrum and included psychos* OR psychotic OR schizophren* OR "at risk mental state" OR "ultra high risk" OR paranoi* OR delusion* OR hallucination*. The Boolean operator AND was used to combine the three primary search term categories.

Initially, the titles and abstracts of identified papers were screened by the second author (FA) against the eligibility criteria to remove duplicates. Subsequently, full-text papers for the remaining papers were obtained where the title and abstract were jointly reviewed by all three authors. Any areas of disagreement between the reviewers about a decision to include or exclude were resolved through discussion. Figure 1 presents the study selection process.

Assessment of Methodological Quality

The methodological quality of eligible studies was reviewed using the Quality Assessment Tool for Quantitative Studies (QATO) assessment tool [42]. The tool, which has good content and construct validity [43,44], was designed to evaluate the quality of quantitative studies in 6 key domains comprising study design, data collection methods, blinding, selection bias, confounding variables, and withdrawals and dropouts.

Each study was assigned a global rating of methodological quality denoting the overall strength of the ratings across the individual domains. Studies with no weak domain ratings were classified as "strong"; those with at least 1 weak rating were classified as "moderate"; and those with ≥2 weak ratings as "weak." All papers were rated by author FA, and 50% of papers were rated by at least two authors (FA, LRV, and JO), with any rating discrepancy resolved through discussion.

Results

Information Extraction

Our database and manual search methods yielded 11,327 papers, which were subsequently reduced to 8538 studies, following the exclusion of duplicates.

Next, we retrieved 28 full-text papers, which were read in full and assessed against the inclusion or exclusion criteria. Overall, we excluded 19 papers subsequently, resulting in a total of 9 studies, derived from 8 unique datasets, which met full criteria for the inclusion in this review and were assessed against the inclusion criteria by all three authors (FA, JO, and LRV). One paper [45] presents follow-up data from their original study [46]. Figure 1 presents the flowchart detailing the extraction.

Qualitative Synthesis

In addition to author details, we extracted relevant characteristics from the selected studies on study origin, rationale, design, sample, and details of the intervention and outcomes (Tables 1 and 2).

Study Characteristics

Study Origin

Regarding the study origin, 4 of the 9 studies were undertaken solely in the United States [45,46,48,52]. Except for one study, the remaining studies were conducted in Hong Kong [47], Ireland [49], Turkey [50], and the United Kingdom [53]. In one study [51], 26 of 30 participants were recruited from the United States and the remaining sample had been recruited from Peru, Australia, and Canada.

Participant Characteristics

Across the studies, the total number of caregiver participants was 305. The individual number of participants reported in each study ranged from 16 to 91 [47,52]. Participants were predominately females in 8 of 9 nine studies that provided gender data [45-51,53]. While more than half the studies employed caregiver participants only, 4 papers reported data from caregiver participants and individuals they provided informal care for [45,46,48,52].

Study Rationale

There was heterogeneity in the reported study objectives, with two-thirds reporting to be an investigation of the feasibility, acceptability, and usability of the intervention [45-48,52,53]. Two studies specifically sought to test the effectiveness of their intervention [49,50]. In the final paper [51], an online self-help group sought to examine the functionality of the user communications.

Table 1. Summary of reviewed studies.

Reference	Origin	Sample	N	Caregiver (%)/gender	QATO[a] rating	Digital intervention type
Chan et al [47]	Hong Kong	First-episode psychosis caregivers	81	75/F[b]	Weak	Website
Glynn et al [48]	United States	Patients with schizophrenia or schizoaffective disorder living in the community and their relatives	42	83/F	Moderate	Website
Haley et al [49]	Ireland	Relatives of people with a schizophrenia spectrum disorder	56	Not reported	Weak	Telepsychiatry videoconferencing
Ozkan et al [50]	Turkey	Primary caregivers of hospitalized patients with schizophrenia	62	53/F	Moderate	Telepsychiatriy telephone
Perron [51]	United States, United Kingdom, Peru, Australia	Relatives of people with schizophrenia and related mental health problems	33	79/F	Weak	Internet: email and bulletin board
Rotondi et al [46]	United States	Persons with schizophrenia spectrum (n=30) and their informal caregivers	21	68/F	Weak	Website
Rotondi et al [45]	United States	Persons with schizophrenia or schizoaffective disorders (n=31) and their informal caregivers	24	63/F	Weak	Website
Ruskin et al [52]	United States	Caregivers of n=22 patients with schizophrenia or schizoaffective disorder	16	Not reported	Weak	eMonitor
Sin et al [53]	United Kingdom	Siblings of 18 people with psychosis	19	84/F	Weak	Website

[a]QATO: Quality Assessment Tool for Quantitative Studies.

[b]F: female.

Table 2. Study aims, intervention components and findings of reviewed studies.

Reference	Study aim	Key intervention components	Main findings
Chan et al [47]	Usability of an internet-based Psychosis Education Program designed to provide up-to-date and interactive online information about psychosis and local resources	• Information about psychosis, caregiver coping, and support • Information on local resources • Downloadable video and written information • Interactive discussion with caregiver peers and professionals	(1) On average, participants used website 2-5 times; (2) 85.2% reported improved knowledge about psychosis; (3) 74.7% felt supported by the site; (4) >80% would recommend the website to others; (5) 80% felt website was easy to use; (6) 81.5% felt website had sufficient information
Glynn et al [48]	Feasibility and quasi-experimental 12-month trial of the online multifamily group program for relatives of persons with schizophrenia	• Discussion board • Resources links • Psychoeducational videos and information • Interactive live chat between participants and professionals on Sunday evenings to focus on problem solving, illness management concerns • Optional groups focused on • Medication • Social support	(1) 79% of caregivers completed the intervention; (2) 52.6% attended core Sunday evening sessions. 84.6% used the discussion board; (3) 30% engaged with optional groups; (4) 92% expressed satisfaction with the intervention; (5)14% reported having initial difficulty with the website; trend significance for patients from experimental group to be hospitalized less during the year of the intervention (24% vs 50%, X^2_1=2.9, P<.09)
Haley et al [49]	To evaluate the effectiveness of caregiver psychoeducation course delivered via telepsychiatry	• 6x2-h educational sessions • Sessions delivered by interactive videoconferencing equipment	Significant increases at postintervention in caregiver knowledge about psychosis
Ozkan et al [50]	Randomized controlled trial to assess the impact of psychoeducation in a hospital clinic and telepsychiatric follow-up after inpatient discharge	• Initial 8 sessions of face-to-face psychoeducation followed by 6 months of regular 15-min phone calls from clinician on set days to facilitate the expression of emotion	Significant postintervention, reduction in the experimental group in levels of caregiver burden, expressed emotion, and depression (P<.001)
Perron [51]	To examine an online self-help group for caregivers of people with mental health problems specifically the patterns and functions of their communications	• An open asynchronous group comprising email and bulletin board	Participants posted an average of 12.6 messages (range, 1-92); male participants posted an average of 1.8 messages versus females who posted an average of 15.5; disclosure (eg, about their participant's lives, their emotions, their relative's condition) was the most common type of message function
Rotondi et al [46]	Randomized controlled trial evaluating the feasibility of random allocation to Schizophrenia Online Access to Resources website intervention delivering online multifamily therapy for persons with schizophrenia and caregivers; 3-month outcomes	• Three professionally led facilitated therapy forums that were for patients only, caregivers only, and patients and caregivers together • Therapy forums focused on problem solving, stress management, and peer support • Ask the "Expert" questions • Educational resources library, including information on local events and relevant news	(1) Therapy groups were the most used component of the website by caregivers; (2) no significant differences in outcomes between caregivers in experimental and treatment as usual groups but patients reported significantly less perceived stress; (3) 27.3% of caregivers reported loneliness when using the website; (4) caregivers suggested areas for the improvement included the greater provision of medication information and research on treatments, and the inclusion of website areas for caregivers to communicate about nonpsychosis-related issues (eg, cooking recipes)
Rotondi et al [45]	Examine use and benefits following random allocation to the Schizophrenia online access to resources website, delivering online multifamily therapy; 12-month outcomes	• Three professionally led facilitated therapy forums that were for patients only, caregivers only, and patients and caregivers together • Therapy forums focused on problem solving, stress management, and peer support. • Ask the "Expert" questions • Educational resources library, including information on local events and relevant news	(1) 92% were engaged in the treatment program; (2) caregivers spent an average of 14 h on the site (range, 30-4021 min) and were in active therapy for an average of 9 months (range, 1-19); (3) significant improvement in caregivers' knowledge of psychosis, specifically, beliefs about prognosis

Reference	Study aim	Key intervention components	Main findings
Ruskin et al [52]	Feasibility of using a home-based computerized device (Med-eMonitor) to enhance the monitoring of patient medication compliance and symptoms and to provide psychoeducation over a 2-month period	• Med-eMonitor records the date and time of when patient medication containers are opened and records when medications are missed • Prompts medication compliance through emitting an audible tone • Liquid crystal display screen that provides factual information on psychosis and poses questions to patients and caregivers about patient clinical status	(1) Caregivers reported significant improvement in knowledge about psychosis (t=2.39, P=.05) but no improvements noted in the patient sample; (2) caregivers believed the monitor impacted positively on patients taking their medication more regularly and helped the caregiver remember to give medication; (3) caregivers reported satisfaction with the monitor though most would opt not to use the monitor after the end of the study
Sin et al [53]	Evaluate user satisfaction and usability over a 4-week period of an online psychoeducation intervention for siblings of people with psychosis "E siblings"	• Downloadable factual information about psychosis • Ask the Experts • Direct access to advice from 12 professionals • Interactive modules on self-care and coping • Frequently Answered Questions • Links to resources • Discussion forum and blog	(1)17 participants completed the full evaluation; (2) participants each spent approximately 2 h (SD 72 min) using the site. Average site visits were 25 mins (SD 12); (3) all participants rated intervention highly, and approximately 95% rated content as very relevant to them; (4) 88.2% rated the intervention as being helpful; (5) 70.5% would recommend the site to others

Length of Intervention

The duration of the intervention under review was not always clear. No information was offered in Chan et al's study [47], while others described interventions that lasted from 1 month [53] to 12 months [45,48]. Comparisons among studies were difficult because the amount of exposure to the given intervention was often evaluated and reported in different ways. While some papers provided data on the total number of minutes a participant was engaged in the intervention [45,47,53], others described the percentage of the sample who engaged with different components of the intervention [48].

Intervention Types

Variability was present in the digital interventions described and their intended outcomes. While 5 studies could be best described as Web-based [45-48,53], 2 studies used a telepsychiatry approach comprising a phone [50] or videoconferencing [49]. The remaining two studies used an eMonitor [52] or combined email and bulletin board [51].

Web-Based

In two of the Web-based studies [47,53], the intervention focused on the provision of information and tailored resources to support caregivers with more informed understanding of psychosis. Chan et al [47] reported data on perceptions of the usefulness and ease of use from 81 participants who accessed an internet-based psychoeducation program for first-episode psychosis caregivers in Hong Kong. The program sought to provide relevant and updated information using downloadable text-based papers, talking head videos (ie, where someone talks directly to the camera) from experts on different aspects of psychosis (eg, cause, treatments, and relapse), caregiving (eg, coping strategies and self-care), and local resources (eg, residential care services and financial services). Information was available in large font size and delivered in English and Cantonese to appeal to and address the needs of a broad group of users. Moreover, the Web platform provided a Web-based and interactive forum, which was moderated and designed to support discussion among peers and between peers and clinicians. Caregivers could post questions and receive responses from clinicians and other caregivers.

Sin et al [53] assessed the usability and feasibility for 16 participants accessing a Web-based platform exclusively dedicated to addressing the information and support needs of siblings of people with psychosis. The platform included 4 main components that focused on the information provision about the illness, coping and well-being, peer discussion forums and blogs, and an "ask the (clinical) expert" feature. The latter allowed participants to post questions to dedicated health care professionals (eg, general practitioner, mental health nurse, and psychiatrist) and receive tailored responses.

The remaining 3 papers, which described 2 studies [45,46,48], used a Web-based platform to deliver family intervention therapies that hitherto were typically implemented through face-to-face meetings held in clinics. Glynn et al [48] completed a proof-of-concept open trial, using a quasi-experimental design, to evaluate the feasibility of a 12-month online multifamily group intervention designed to provide education and support.

In their study, 26 caregivers of community-dwelling patients were involved in the intervention and data were compared with a historical treatment as usual sample (n=16). Caregivers had to have access to computers at their home to be eligible to participate; the Web program also comprised a discussion board, links to relevant resources, and educational videos. Participants were organized into small groups of 5-6 caregiver participants. During 1 year (comprising 6 months on a weekly basis, and biweekly for further 6 months), participants could access an hour-long educational talk on problem solving, and goal setting sessions with a psychologist and a research staff member. These sessions were held on Sunday evenings. Furthermore, caregivers could access additional groups on medication and support. The authors noted their small sample size despite "intensive" recruitment strategies implemented. That said, participants

reported high levels of satisfaction (>90%) with the intervention, with 84.6% engaging with the discussion board, 52.6% with the Sunday talk sessions, and one-third attending the additional groups. No impact of the intervention was observed on the levels of caregivers' reported distress (or patient functioning). There was a trend of significance that the intervention was linked to fewer patient admissions during the intervention year.

Rotondi et al [45,46] completed a randomized feasibility trial of multifamily psychoeducational interventions also using a Web platform. Unlike Glynn et al [48], in this study, participants were issued computers in their home, if required. In addition, 21 caregivers of inpatients and community-dwelling patients, with a history of at least 1 hospital admission in the preceding 2 years were randomly assigned to the Web intervention or treatment as usual. In the intervention arm, caregivers were issued with a unique log-in name (which was not allowed to be their real name) and password access to a caregiver-only website therapy group and a joint group designed for caregivers and patients together. The therapy group focused on problem solving and offered a bulletin board for communication among group members. The groups were led by mental health professionals and guided by therapy manuals [54,55]. As part of additional intervention modules, caregivers were given opportunities to anonymously pose questions to experts, to receive responses to their questions, and to view questions asked by other caregivers and responses they received. Moreover, they had access to relevant reading material and local relevant mental health news. Before commencing the Web-based intervention, all participants (caregivers and patients) were required to attend a joint 4-hour psychoeducation workshop. Furthermore, outcome data were collected at 3, 6, and 12 months. The most used intervention components within the platform were the two therapy groups (ie, caregiver-only group and caregiver and patient group). The authors failed to identify any difference in outcomes between caregivers in the experimental and treatment arms; however, the patient group reported markedly less stress. Nearly one-third of caregivers reported feeling lonely in their use of the website.

Telepsychiatry

In terms of the two telepsychiatry studies, both focused on the provision of psychoeducation to improve caregivers' knowledge and understanding of the illness and to promote more effective coping strategies. Haley et al [49] delivered a psychoeducation course to 56 caregivers in Dublin and Donegal regions of Ireland using an interactive videoconferencing system that included a Tandberg Director, camera, plasma monitors, and three Integrated Services Digital Network lines. Participants recorded a marked increment in their knowledge postintervention. Ozkan et al [50] used a randomized controlled design to evaluate the impact of providing short (ie, 15 minutes) telephone calls over a 6-month period to caregivers. Caregivers received these calls following their relative's discharge from hospital and following their (ie, caregiver) own participation in an 8-session, face-to-face psychoeducation intervention during the admission. The calls were designed to focus on caregivers' emotions and their experience of burden. The intervention group recorded markedly lower rates of caregiver burden, depression, and emotional expression.

Email or Bulletin Board

Ruskin et al [52] assessed the feasibility of using a computer-based device "Med-eMonitor," which could be preprogrammed to ask questions and display illness-related factual information (eg, schizophrenia incidence rates), to assist patients in improving their daily medication adherence and provide caregivers with psychoeducation. The study duration was 2 months, and 16 caregivers used the monitor to get responses to questions about how they understood psychosis (eg, symptoms). Marked improvements were noted in caregivers' understanding about the illness and their ability to remind their relatives to take their medications. However, they preferred not to use the monitor once the study period ended.

Perron [51] presented data from 33 participants who engaged in an online self-help group for mental health caregivers over an 18-month period. It was an open group (ie, free and open to the general public to engage with), with no moderator and organized around email exchanges and bulletin boards. Perron [51] analyzed the content of 417 messages posted within 18 months. Participants posted an average of 12.6 messages. Male participants posted fewer messages than with female participants (average, 1.8 vs 15.5). The most common category for a posted message was "disclosure." These were messages where participants tended to provide updates on their lives, their experiences with their relative (eg, treatment-related issues and symptoms), and their own emotions. Other key functions identified from posted messages focused on the provision of information, support, and empathy to caregiver peers. Only one message's content was rated as being negative, which did not receive any replies or comments from other caregivers.

Methodological Quality

Overall, 7 papers were scored weakly on the overall QATO rating; these had weakness in the study design, approaches to data collection, management of confounding variables, and assessor independence. In addition, 2 studies [48,50] obtained an overall QATO rating of moderate and had a stronger study design (eg, randomized controlled trial or quasi-experimental) and more robust participant selection and data collection approaches (Table 1).

Discussion

Principal Findings

The literature highlights an increasing interest in and documentation of digital technologies in health care and their novel applications with psychosis disorders. Improved patient outcomes in psychosis rely heavily on the input from informal caregivers. This systematic review explored the application and outcomes of digital technologies in informal caregivers of people with psychosis.

The review yielded 9 papers reporting data from 8 independent studies. Overall, two-thirds of the included studies focused solely on the recruitment and outcomes of caregivers only and one-third recruited caregiver and patient dyads. The reviewed studies were diverse, with origins in Europe, Central and North America, Asia, and the Middle East, in reflecting the global developments in digital technologies. However, most studies

were from the United States, which might reflect different health care priorities that predominate in that region and broader developments in digital innovations. As an estimated 89% of the US population accesses the internet [56], evaluating the contribution of digital interventions to improving outcomes in caregiver groups might seem a sensible development.

There was also diversity in the digital applications under review. The interventions included telepsychiatry and email. Most studies were, however, best described as being Web-based and included those with a dedicated platform with functionality to facilitate communication between caregiver groups and caregivers and professionals. Remarkably, no study identified detailing the use of mobile apps and virtual or augmented realities, which contrasts markedly with developments the literature has observed in psychosis patient populations [57,58]. There is no evidence suggesting that caregivers would be any less likely than peers and other groups to take advantage of mobile phone apps or relevant health-focused augmented realities. A general commitment to exploring and investing in opportunities to expand the range of digital approaches on offer or applicable to caregivers should be prioritized to minimize gaps in service provision and the potential for a digital divide between caregivers and others.

The review findings indicate that we are far from being in a position to offer definitive data about the use and impact of digital innovations in caregivers. The majority of studies under review were described from the outset as being studies of the feasibility and usability. Though studies provided useful data, in the absence of powered experimental designs, definitive conclusions about outcomes and effects are premature.

In the majority of studies, the reason underlying the development and use of the digital tool was the provision of relevant information about psychosis (psychoeducation), delivered as part of a structured educational course or through different independent and related modules. A key component in treatment recommendations for caregivers of people with psychosis [23] is information on how to best understand the illness and facilitate the use of adaptive and effective coping strategies. Psychoeducation is a need commonly reported by caregivers [59], but an area that is often unmet by service providers [60,61]. Over the course of the illness, caregivers will often be expected to make sense of illness-related information that can at times be complex, confusing, and vague. For many, this information will be given during periods when they are also experiencing high levels of stress and therefore perhaps more likely to benefit from varied methods of sharing the relevant information. Though preliminary, the early indications from the findings suggest that caregivers' understanding about psychosis might benefit from the use of digital technologies and that the approach might be acceptable [45,47,49]. Further exploration of the benefits of using different approaches to support caregivers in facilitating their knowledge and understanding of their relative's mental health condition, which also extends beyond traditional face-to-face meetings [62], would represent important research developments and progress in the field.

We do, however, remain aware that at least one-third of caregivers in one website study reported feelings of loneliness in using the intervention [46], and in another study [52], caregivers would opt not to continue using the digital equipment when the study ended. However, at this stage, it remains unclear as to what extent these types of findings generalize and form a distinct pattern. Hence, further evaluations utilizing quantitative and qualitative investigations are indicated. Though the interest in and appetite for digital technologies in mental health sectors remains on a steady upward trend, these findings also underscore the importance of seeking to identify caregiver subgroups for whom digital approaches might not always suit and to address their specific presenting needs.

This study suggests a lack of uniformity in terms of the key areas to measure as outcomes. It is also noteworthy that the website and email or bulletin intervention studies included components that promoted and allowed for peer interactions and support. However, measures of social support and social networks were absent. The current evidence attests that psychosis caregivers are up to 10 times more socially isolated than the general population and typically fair worse in terms of support levels when compared with caregivers of adults with similar challenging conditions [63]. It is important to extend the digital technology outcome literature beyond the rates of take-up and satisfaction. The results indicate a need for further work to be undertaken to identify target outcome areas for measurement in the use of digital technologies with caregivers and the preferred methods of assessment, which in turn could lead to more meaningful evaluation of studies and comparison of findings.

Notably, however, there were 2 studies that evaluated a telepsychiatry [50] and website-based intervention [48] that obtained overall moderate ratings; this reflected their superiority in the study design (eg, randomized controlled trial or quasi-experimental) and participant selection and data collection approaches.

Limitations

The focus of the review on caregivers of people with psychosis is a strength of this study and is in line with other reviews focusing on technology in different caregiver populations such as older adults [64] and severe mental illnesses [65]. The review, however, does have some limitations. First, the studies were mainly investigations of feasibility, usability, and acceptability and were therefore not designed or powered to offer definitive conclusions about the efficacy and direction of findings, which largely reflects the poor methodological quality ratings. The majority of papers (n=7) obtained weak QATO ratings that were representative of a broad range of issues reflecting inherent difficulties in the study design, approaches to data collection, management of confounding variables, and assessor independence. However, given that some might argue about the relative infancy of the literature and the predominance of study designs (eg, usability and feasibility), the ratings might simply reflect the stage of the literature with stronger and more methodologically sound studies to follow.

Second, a modest number of studies were under review and the overall participant size was small. The smaller studies reported samples of n=16. Notably, the recruitment pathways for some studies were dependent on both patient and caregiver consenting

to participate [48], while others could directly recruit and consent caregivers [53]. The different recruitment approaches across studies are likely to have implications for the sample and their presenting needs. For example, studies that required patient and caregiver consent might be more likely to recruit groups who were better functioning or had better quality caregiving relationships, which arguably is likely to impact their engagement. The majority of studies were from the United States, a high-income nation with large sections of the adult population accessing the internet. The review was limited to English language publications. Consequently, the interpretation of findings and their generalizability to other settings and communities are limited. Furthermore, while we may have sought to be overinclusive in our search approach, it remains possible that we might have missed potentially relevant studies, given the language restrictions and parallel exclusions of case studies and qualitative investigations. Not dissimilar to other systematic reviews, it is possible that our review will be subject to publication bias because nonsignificant findings are less likely to be published. Therefore, the reviewed studies could overrepresent the positive effects of digital interventions with caregivers.

Conclusions

Notwithstanding the continued value of direct service input and face-to-face contact, the potential contribution of digital interventions to impacting outcomes for psychosis caregivers and addressing their specific needs for information, support, and well-being deserves greater clinical and research interest. Evidently, given the number and range of studies reviewed, widely established digital developments witnessed in patient mental health care have yet to be replicated in caregiver populations. However, this study offers preliminary support that these types of interventions (eg, Web-based) can be feasible and acceptable to caregivers. Much further development in the range of technologies on offer and robust evaluation of their outcomes, including cost-effectiveness [66], is required. Furthermore, the study indicates the inclusion of caregiver-reported outcomes and their qualitative reports of satisfaction.

Conflicts of Interest

None declared.

References

1. Fleischhacker W, Arango C, Arteel P, Barnes T, Carpenter W, Duckworth K, et al. Schizophrenia--time to commit to policy change. Schizophr Bull 2014 Apr;40 Suppl 3:S165-S194 [FREE Full text] [doi: 10.1093/schbul/sbu006] [Medline: 24778411]
2. Hasan A, Falkai P, Wobrock T, Lieberman J, Glenthoj B, Gattaz WF, World Federation of Societies of Biological Psychiatry (WFSBP) Task Force on Treatment Guidelines for Schizophrenia. World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Schizophrenia, part 1: update 2012 on the acute treatment of schizophrenia and the management of treatment resistance. World J Biol Psychiatry 2012 Jul;13(5):318-378. [doi: 10.3109/15622975.2012.696143] [Medline: 22834451]
3. Kuipers E, Bebbington P, Dunn G, Fowler D, Freeman D, Watson P, et al. Influence of carer expressed emotion and affect on relapse in non-affective psychosis. Br J Psychiatry 2006 Feb;188:173-179. [doi: 10.1192/bjp.bp.104.007294] [Medline: 16449706]
4. Sadath A, Muralidhar D, Varambally S, Gangadhar BN, Jose JP. Do stress and support matter for caring? The role of perceived stress and social support on expressed emotion of carers of persons with first episode psychosis. Asian J Psychiatr 2017 Feb;25:163-168. [doi: 10.1016/j.ajp.2016.10.023] [Medline: 28262142]
5. Revier CJ, Reininghaus U, Dutta R, Fearon P, Murray RM, Doody GA, et al. Ten-Year Outcomes of First-Episode Psychoses in the MRC ÆSOP-10 Study. J Nerv Ment Dis 2015 May;203(5):379-386 [FREE Full text] [doi: 10.1097/NMD.0000000000000295] [Medline: 25900547]
6. Norman RMG, Malla AK, Manchanda R, Harricharan R, Takhar J, Northcott S. Social support and three-year symptom and admission outcomes for first episode psychosis. Schizophr Res 2005 Dec 15;80(2-3):227-234. [doi: 10.1016/j.schres.2005.05.006] [Medline: 15964175]
7. Stowkowy J, Addington D, Liu L, Hollowell B, Addington J. Predictors of disengagement from treatment in an early psychosis program. Schizophr Res 2012 Apr;136(1-3):7-12. [doi: 10.1016/j.schres.2012.01.027] [Medline: 22336955]
8. Garety PA, Fowler DG, Freeman D, Bebbington P, Dunn G, Kuipers E. Cognitive--behavioural therapy and family intervention for relapse prevention and symptom reduction in psychosis: randomised controlled trial. Br J Psychiatry 2008 Jun;192(6):412-423. [doi: 10.1192/bjp.bp.107.043570] [Medline: 18515890]
9. The SC. The abandoned illness: a report from the Schizophrenia Commission. London: Rethink Mental Illness; 2012:a.
10. Poon A, Harvey C, Mackinnon A, Joubert L. Longitudinal population-based study of carers of people with psychosis. Epidemiol Psychiatr Sci 2016;5:1-11.
11. Caring for carers survey. URL: http://www.eufami.org/c4c/ [accessed 2018-08-06] [WebCite Cache ID 71Ta6K55l]
12. Kuipers E, Raune D. The early development of expressed emotionburden in the families of first onset psychosis. M. Birchwood, D Fowler, C Jackson. Eds. Chichester: Wiley 2000.

13. Hamaie Y, Ohmuro N, Katsura M, Obara C, Kikuchi T, Ito F, et al. Criticism and Depression among the Caregivers of At-Risk Mental State and First-Episode Psychosis Patients. PLoS One 2016;11(2):e0149875 [FREE Full text] [doi: 10.1371/journal.pone.0149875] [Medline: 26918629]
14. Patterson P, Birchwood M, Cochrane R. Expressed emotion as an adaptation to loss Prospective study in first-episode psychosis. Br J Psychiatry 2005;187(48):59-64. [doi: 10.1192/bjp.187.48.s59]
15. Perlick D, Hohenstein J, Clarkin J, Kaczynski R, Rosenheck R. Use of mental health and primary care services by caregivers of patients with bipolar disorder: a preliminary study. Bipolar Disorders 2005;7:a. [doi: 10.1111/j.1399-5618.2004.00172.x]
16. Gupta S, Isherwood G, Jones K, Van IK. Assessing health status in informal schizophrenia caregivers compared with health status in non-caregivers and caregivers of other conditions. BMC Psychiatry 2015 Jul 21;15:162 [FREE Full text] [doi: 10.1186/s12888-015-0547-1] [Medline: 26194890]
17. Smith L, Onwumere J, Craig T, Kuipers E. A role for poor sleep in determining distress in caregivers of individuals with early psychosis. Early Interv Psychiatry. In press 2018 (forthcoming).
18. Claxton M, Onwumere J, Fornells-Ambrojo M. Do family interventions improve outcomes in early psychosis? a systematic review and meta-analysis? Frontiers in Psychology 2017; 8. Doi.org/10.3389/fpsyg.0 2017:0371. [doi: 10.3389/fpsyg.2017.00371]
19. Pharoah F, Mari J, Rathbone J, Wong W. Family intervention for schizophrenia. Cochrane Database Syst Rev 2010 Dec 08(12):CD000088 [FREE Full text] [doi: 10.1002/14651858.CD000088.pub2] [Medline: 21154340]
20. Norman R, Lecomte T, Addington D, Anderson E. Canadian Treatment Guidelines on Psychosocial Treatment of Schizophrenia in Adults. Can J Psychiatry 2017 Sep;62(9):617-623. [doi: 10.1177/0706743717719894] [Medline: 28703017]
21. Galletly C, Castle D, Dark F, Humberstone V, Jablensky A, Killackey E, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the management of schizophrenia and related disorders. Aust N Z J Psychiatry 2016 May;50(5):410-472. [doi: 10.1177/0004867416641195] [Medline: 27106681]
22. Kreyenbuhl J, Buchanan RW, Dickerson FB, Dixon LB, Schizophrenia Patient Outcomes Research Team (PORT). The Schizophrenia Patient Outcomes Research Team (PORT): updated treatment recommendations 2009. Schizophr Bull 2010 Jan;36(1):94-103 [FREE Full text] [doi: 10.1093/schbul/sbp130] [Medline: 19955388]
23. National Institute for Health and Care Excellence. Psychosis and Schizophrenia in adults: treatment and management. In: Clinical Guideline 178. London: NICE; 2014.
24. Ince P, Haddock G, Tai S. A systematic review of the implementation of recommended psychological interventions for schizophrenia: Rates, barriers, and improvement strategies. Psychol Psychother 2016 Dec;89(3):324-350. [doi: 10.1111/papt.12084] [Medline: 26537838]
25. Onwumere J, Grice S, Kuipers E. Delivering Cognitive-Behavioural Family Interventions for Schizophrenia. Australian Psychologist 2016 Jan 20;51(1):52-61. [doi: 10.1111/ap.12179]
26. Eassom E, Giacco D, Dirik A, Priebe S. Implementing family involvement in the treatment of patients with psychosis: a systematic review of facilitating and hindering factors. BMJ Open 2014 Oct 03;4(10):e006108 [FREE Full text] [doi: 10.1136/bmjopen-2014-006108] [Medline: 25280809]
27. Internet Live Stats. Internet Users URL: http://www.internetlivestats.com/internet-users/ [accessed 2018-07-31] [WebCite Cache ID 71Jc2aHDk]
28. Office for National Statistics. Internet access – households and individuals: 2017 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017 [accessed 2018-07-31] [WebCite Cache ID 71Jc83lkv]
29. Princess Royal Trust for Carers. How can the web support carers URL: https://carers.org/section/campaigns [accessed 2018-08-06] [WebCite Cache ID 71Ta9Q8q4]
30. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
31. Biagianti B, Hidalgo-Mazzei D, Meyer N. Developing digital interventions for people living with serious mental illness: perspectives from three mHealth studies. Evid Based Mental Health 2017;20(2):2765. [doi: 10.1136/eb-2017-102765]
32. Elfeddali I, Bolman C, Candel M, Wiers R, de VH. Preventing smoking relapse via Web-based computer-tailored feedback: a randomized controlled trial. J Med Internet Res 2012 Aug 20;14(4):e109 [FREE Full text] [doi: 10.2196/jmir.2057] [Medline: 22903145]
33. Cunningham JA, Gulliver A, Farrer L, Bennett K, Carron-Arthur B. Internet interventions for mental health and addictions: current findings and future directions. Curr Psychiatry Rep 2014 Dec;16(12):521. [doi: 10.1007/s11920-014-0521-5] [Medline: 25308390]
34. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015 May;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
35. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]

36. Arnberg F, Linton S, 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. [doi: 10.1371/journal.pone.0098118]
37. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
38. Reeder C, Pile V, Crawford P, Cella M, Rose D, Wykes T, et al. The Feasibility and Acceptability to Service Users of CIRCuiTS, a Computerized Cognitive Remediation Therapy Programme for Schizophrenia. Behav Cogn Psychother 2016 May;44(3):288-305 [FREE Full text] [doi: 10.1017/S1352465815000168] [Medline: 26004421]
39. Cella M, Okruszek L, Lawrence M, Zarlenga V, He Z, Wykes T. Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. Schizophr Res 2017 Oct 03. [doi: 10.1016/j.schres.2017.09.028] [Medline: 28986005]
40. Naslund J, Aschbrenner K, Bartels SJ. Wearable Devices and Smartphones for Activity Tracking Among People with Serious Mental Illness. Ment Health Phys Act 2016 Mar;10:10-17 [FREE Full text] [doi: 10.1016/j.mhpa.2016.02.001] [Medline: 27134654]
41. Moher D. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Intern Med 2009 Aug 18;151(4):264. [doi: 10.7326/0003-4819-151-4-200908180-00135]
42. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x] [Medline: 17163895]
43. Armijo-Olivo S, Stiles C, Hagen N, Biondo P, Cummings G. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment tool: methodological research. J Eval Clin Pract 2012;18(1):a. [doi: 10.1111/j.1365-2753.2010.01516.x]
44. Jackson N, Waters E, Guidelines for Systematic Reviews in Health PromotionPublic Health Taskforce. Criteria for the systematic review of health promotion and public health interventions. Health Promot Int 2005 Dec;20(4):367-374. [doi: 10.1093/heapro/dai022] [Medline: 16169885]
45. Rotondi A, Anderson C, Haas G, Eack S, Spring M, Ganguli R. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010;61(11):1099-1105. [doi: 10.1176/appi.ps.61.11.109]
46. Rotondi AJ, Haas GL, Anderson CM, Newhill CE, Spring MB, Ganguli R, et al. A Clinical Trial to Test the Feasibility of a Telehealth Psychoeducational Intervention for Persons With Schizophrenia and Their Families: Intervention and 3-Month Findings. Rehabil Psychol 2005 Nov;50(4):325-336 [FREE Full text] [doi: 10.1037/0090-5550.50.4.325] [Medline: 26321774]
47. Chan S, Tse S, Sit HLT, Hui C, Lee E, Chang W, et al. Web-Based Psychoeducation Program for Caregivers of First-Episode of Psychosis: An Experience of Chinese Population in Hong Kong. Front Psychol 2016;7:2006 [FREE Full text] [doi: 10.3389/fpsyg.2016.02006] [Medline: 28082935]
48. Glynn SM, Randolph ET, Garrick T, Lui A. A proof of concept trial of an online psychoeducational program for relatives of both veterans and civilians living with schizophrenia. Psychiatr Rehabil J 2010;33(4):278-287. [doi: 10.2975/33.4.2010.278.287] [Medline: 20374986]
49. Haley C, O'Callaghan E, Hill S, Mannion N, Donnelly B, Kinsella A, et al. Telepsychiatry and carer education for schizophrenia. Eur Psychiatry 2011 Jul;26(5):302-304. [doi: 10.1016/j.eurpsy.2009.12.021] [Medline: 20542666]
50. Ozkan B, Erdem E, Demirel OS, Zararsiz G. Effect of psychoeducation and telepsychiatric follow up given to the caregiver of the schizophrenic patient on family burden, depression and expression of emotion. Pak J Med Sci 2013 Sep;29(5):1122-1127 [FREE Full text] [Medline: 24353704]
51. Perron B. Online support for caregivers of people with a mental illness. Psychiatr Rehabil J 2002;26(1):70-77. [Medline: 12171285]
52. Ruskin P, Van Der Wende J, Clark C, Fenton J, Deveau J, Thapar R, et al. Feasibility of Using the Med-eMonitor System in the Treatment of Schizophrenia: A Pilot Study. Drug Information Journal 2016 Aug 28;37(3):283-291. [doi: 10.1177/009286150303700304]
53. Sin J, Henderson C, Norman I. Usability of online psychoeducation for siblings of people with psychosis. Int J Technol Assess Health Care 2014 Oct;30(4):374-380. [doi: 10.1017/S0266462314000488] [Medline: 25394550]
54. Anderson C, Reiss D, Hogarty G. Schizophrenia and the family: A practitioner's guide to psychoeducation and management. New York: Guilford Press; 1986.
55. McFarlane W, Deakins S, Gingerich S, Dunne E, Horen B, Newmark M. Multiple-Family Psychoeducational Group Treatment Manual. New York: Biosocial Treatment Division, New York State Psychiatric Institute; 1991.
56. Pew Research Center. Internet/Broadband Fact Sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2018-08-06] [WebCite Cache ID 71oDtKUvp]
57. Valmaggia L. The use of virtual reality in psychosis research and treatment. World Psychiatry 2017 Oct;16(3):246-247 [FREE Full text] [doi: 10.1002/wps.20443] [Medline: 28941114]

58. Valmaggia L, Latif L, Kempton M, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: a systematic review of recent evidence. Psychiatry Res 2016;236:189-195. [doi: 10.1016/j.psychres.2016.01.015]
59. Askey R, Holmshaw J, Gamble C. Gray R. What do carers of people with psychosis need from mental health services? Exploring the views of carers, service users and professionals. J Fam Ther 2009;31:310-331. [doi: 10.1111/j.1467-6427.2009.00470.x]
60. Giacco D, Dirik A, Kaselionyte J, Priebe S. How to make carer involvement in mental health inpatient units happen: a focus group study with patients, carers and clinicians. BMC Psychiatry 2017;17:a. [doi: 10.1186/s112888-017-1259-5]
61. Cree L, Brooks HL, Berzins K, Fraser C, Lovell K, Bee P. Carers' experiences of involvement in care planning: a qualitative exploration of the facilitators and barriers to engagement with mental health services. BMC Psychiatry 2015 Aug 29;15:208 [FREE Full text] [doi: 10.1186/s12888-015-0590-y] [Medline: 26319602]
62. Onwumere J, Glover N, Whittaker S, Rahim S, Chu ML, James G, et al. Modifying illness beliefs in recent onset psychosis carers: Evaluating the impact of a cognitively focused brief group intervention in a routine service. Early Interv Psychiatry 2017 May 18. [doi: 10.1111/eip.12430] [Medline: 28517041]
63. Hayes L, Hawthorne G, Farhall J, O'Hanlon B, Harvey C. Quality of Life and Social Isolation Among Caregivers of Adults with Schizophrenia: Policy and Outcomes. Community Ment Health J 2015 Jul;51(5):591-597. [doi: 10.1007/s10597-015-9848-6] [Medline: 25690154]
64. Guay C, Auger C, Demers L, Mortenson WB, Miller WC, Gélinas-Bronsard D, et al. Components and Outcomes of Internet-Based Interventions for Caregivers of Older Adults: Systematic Review. J Med Internet Res 2017 Sep 19;19(9):e313 [FREE Full text] [doi: 10.2196/jmir.7896] [Medline: 28928109]
65. Meyer T, Casarez R, Mohite S, La Rosa RN, Iyengar S. Novel technology as platform for interventions for caregivers and individuals with severe mental health illnesses: A systematic review. J Affect Disord 2018 Jan 15;226:169-177. [doi: 10.1016/j.jad.2017.09.012] [Medline: 28987999]
66. 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]

Abbreviations

QATO: Quality Assessment Tool for Quantitative Studies

Edited by J Torous; submitted 16.01.18; peer-reviewed by S Chan, D Gélinas-Bronsard; comments to author 27.02.18; revised version received 11.04.18; accepted 19.05.18; published 05.09.18.

Please cite as:
Onwumere J, Amaral F, Valmaggia LR
Digital Technology for Caregivers of People With Psychosis: Systematic Review
JMIR Ment Health 2018;5(3):e55
URL: http://mental.jmir.org/2018/3/e55/
doi:10.2196/mental.9857
PMID:30185402

Review

For Better or for Worse? A Systematic Review of the Evidence on Social Media Use and Depression Among Lesbian, Gay, and Bisexual Minorities

César G Escobar-Viera[1,2*], MD, PhD; Darren L Whitfield[2,3,4*], PhD; Charles B Wessel[5*], MLS; Ariel Shensa[1*], MA; Jaime E Sidani[1*], PhD; Andre L Brown[2*], PhD; Cristian J Chandler[2*], PhD, MPH; Beth L Hoffman[1*], BSc; Michael P Marshal[4*], PhD; Brian A Primack[1*], MD, PhD

[1]Center for Research on Media, Technology, and Health, School of Medicine, University of Pittburgh, Pittsburgh, PA, United States

[2]Center for LGBT Health Research, University of Pittsburgh, Pittsburgh, PA, United States

[3]School of Social Work, University of Pittsburgh, Pittsburgh, PA, United States

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

[5]Health Sciences Library, University of Pittsburgh, Pittsburgh, PA, United States

[*]all authors contributed equally

Corresponding Author:
César G Escobar-Viera, MD, PhD
Center for Research on Media, Technology, and Health
School of Medicine
University of Pittburgh
230 McKee Place
Suite 600.11
Pittsburgh, PA, 15213
United States
Phone: 1 412 692 4297
Fax: 1 412 692 4838
Email: escobar-viera@pitt.edu

Abstract

Background: Over 90% of adults in the United States have at least one social media account, and lesbian, gay, and bisexual (LGB) persons are more socially active on social media than heterosexuals. Rates of depression among LGB persons are between 1.5- and 2-fold higher than those among their heterosexual counterparts. Social media allows users to connect, interact, and express ideas, emotions, feelings, and thoughts. Thus, social media use might represent both a protective and a risk factor for depression among LGB persons. Studying the nature of the relationship between social media use and depression among LGB individuals is a necessary step to inform public health interventions for this population.

Objective: The objective of this systematic review was to synthesize and critique the evidence on social media use and depression among LGB populations.

Methods: We conducted a literature search for quantitative and qualitative studies published between January 2003 and June 2017 using 3 electronic databases. Articles were included if they were peer-reviewed, were in English, assessed social media use either quantitatively or qualitatively, measured depression, and focused on LGB populations. A minimum of two authors independently extracted data from each study using an a priori developed abstraction form. We assessed appropriate reporting of studies using the Strengthening the Reporting of Observational Studies in Epidemiology and the Consolidated Criteria for Reporting Qualitative Research for quantitative and qualitative studies, respectively.

Results: We included 11 articles in the review; 9 studies were quantitative and cross-sectional and 2 were qualitative. Appropriate reporting of results varied greatly. Across quantitative studies, we found heterogeneity in how social media use was defined and measured. Cyberbullying was the most studied social media experience and was associated with depression and suicidality. Qualitative studies found that while social media provides a space to disclose minority experiences and share ways to cope and get support, constant surveillance of one's social media profile can become a stressor, potentially leading to depression. In most studies, sexual minority participants were identified inconsistently.

Conclusions: This review supports the need for research on the role of social media use on depression outcomes among LBG persons. Using social media may be both a protective and a risk factor for depression among LGB individuals. Support gained via social media may buffer the impact of geographic isolation and loneliness. Negative experiences such as cyberbullying and other patterns of use may be associated with depression. Future research would benefit from more consistent definitions of both social media use and study populations. Moreover, use of larger samples and accounting for patterns of use and individuals' experiences on social media may help better understand the factors that impact LGB mental health disparities.

(JMIR Ment Health 2018;5(3):e10496) doi:10.2196/10496

KEYWORDS

social media; social networking sites; sexual minorities; lesbian; gay; bisexual; depression; systematic review

Introduction

Despite growing acceptance and civil rights gains in recent years, lesbian, gay, and bisexual (LGB) individuals in the United States still face stigma and disparities regarding mental health conditions [1]. LGB persons are a diverse population whose sexual attraction, behavior, or orientation differs from their heterosexual counterparts. Importantly, estimated rates of depression among LGB persons are between 1.5- and 2-fold higher than their heterosexual counterparts [2]. In 2015, 14.8% of LGB males and 20.4% of LGB females in the United States suffered at least one major depressive episode, totaling 1.9 million individuals, compared with 4.3% of heterosexual males and 8% of heterosexual females [1]. These findings are consistent with those in other developed countries, and disparities are greater among bisexual adults [3,4].

Social media includes a variety of websites and mobile apps that enable users to create content and participate in online social networking (eg, YouTube, Tumblr, Facebook) [5]. It is estimated that well over 90% of adults in the United States have at least one social media account, with an average daily use of 2-4 hours [6,7]. Social media is a communication space where users express emotions, feelings, and thoughts. For LGB individuals, social media is a primary mode of socializing, and LGB persons are more socially active on social media than heterosexuals [8,9]. National data on LGB individuals found that over 85% of participants had one social media account, and they used it at least weekly; this usage rose to over 91% among LGB young adults [10].

Social media use encompasses a series of measures that capture the experience of using social media. Although no clear consensus exists regarding which specific measures should be counted as social media use, common ones include time (time elapsed while using social media over 24 hours), frequency (number of times people check their social media account per day) [11], number of friends and self-presentation [12], number of platforms (sites or apps) used [13], closeness to online friends, and activities performed (eg, posting updates, sharing pictures, etc) [14,15], as well as other use patterns (eg, active vs passive use, experiences such as experiencing cyberbullying, problematic use, positive or negative quality of interactions, and motivation to use social media) [16-22].

In a broader context, previous research has also investigated the linguistic attributes of social media use that indicate self-disclosure and mental health information sharing and that predict receipt of social support and other therapeutic outcomes on social media [23,24]. Moreover, several studies have looked at either the moderating or mediating role of upward social comparison [25-30], social connectedness [31], envy [32], and intensity of use [33].

Some of the complexities associated with social media use among LGB persons have been studied in relation with parenting and gender transition [34,35]. The findings of these studies point out potential stressors, such as the need to become incidental advocates or the task of detecting disapproval and allies within one's social networks [34]. On the other hand, social media has some duality to it; despite being a stressful environment, it can also provide support that helps mitigate the said stress [35]. More specifically, while some researchers have found an association between social media use and increased risk of depression [11,13,16,36], others have found an association between specific patterns of use and improvement of psychological well-being [37]. Thus, social media use may be both a risk and a protective factor for depression and psychological well-being in the general population. In addition, using social media may add unique protective and risk factors for depression among LGB individuals.

Meyer's minority stress theory is the predominant framework for understanding the protective and risk factors for depression and other mental health disparities among LGB individuals [38]. There are at least three fundamental tenets to the minority stress theory. First, exposure to LGB-related stressors such as discrimination, social rejection, and sometimes violent victimization is a central cause of mental health problems among LGB individuals. Second, exposure to these "distal" stressors is associated with "proximal" stressors such as internalized homonegativity and expectations of rejection. Third, social support from within the LGB communities can help offset or buffer the impact of these stressors on mental health outcomes. Meyer's theory has also been extended to incorporate specific groups among LGB persons, such as children, adolescents, and people living with HIV [39,40]. While this theory helps explain the effects of social stress due to marginalized social identities on mental health outcomes among sexual minorities [38], the virtual social environment of social media introduces new complexities to previously described social interactions. For example, social media may make it easier for LGB individuals to disclose their sexual orientation to others by forming connections, providing education, and facilitating positive interactions and social support among LGB individuals. These virtual interactions may reduce the stress experienced by LGB

individuals based on their sexual orientation and may protect sexual minorities from depression [41-43]. This may be particularly true for LGB individuals for whom it is too dangerous to be "out" or disclose their sexual minority identity in real-world settings such as in the workplace or in public social spaces. Conversely, social media use may be a vehicle for negative experiences, such as stigmatization and social comparison. These can lead to negative outcomes, including decreased self-esteem, and depressive symptoms [44,45].

The ubiquity of social media has led researchers to suggest its use to provide people with mental illness the opportunity to challenge stigma, provide and receive peer-to-peer support, and access either Web-based or mobile interventions [46]. To do this, we need to understand the mixed effects of social media use—improvement of psychological well-being or worsening of affective symptoms—on mental health. This understanding will inform policy and studies that leverage the positive aspects and address the potentially negative aspects of use as well. However, to date, no comprehensive synthesis of research on the impact of different patterns of use on depression among LGB populations has been conducted. Considering that LGB persons exhibit consistently high rates of social media use and consistently higher rates of depression than heterosexuals, this is a particularly important gap in the literature. For these reasons, we conducted a systematic review with four overarching goals: (1) identify all the peer-reviewed published papers that examined social media use and depression among LGB individuals; (2) describe the characteristics of these studies, including the study-appropriate reporting and methodology (eg, quantitative vs qualitative); (3) describe how social media use and depression constructs were operationalized across studies; and (4) assess which of the main tenets of the minority stress theory was analyzed in each study, in order to make recommendations for future studies that could leverage social media for improving depression outcomes in this population.

Methods

Inclusion and Exclusion Criteria

This systematic review has been reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement guidelines (Multimedia Appendix 1) [47,48]. The research protocol was registered in the PROSPERO database (#CRD42018088165) and is available as a supplement (Multimedia Appendix 2).

We included quantitative and qualitative studies published in peer-reviewed journals, in the English language, during or after 2003 (when MySpace, the first modern social media site, started operating). We allowed manuscripts from conference proceedings only when full-research papers were required for submission and each submission went through a complete peer-review process. Included studies had to focus on social media use and depression among LGB minorities. We defined social media use as any usage measurement (eg, time, frequency, motivation to use, experiences while using, etc). Depression comprised major depressive disorder, bipolar depression, dysthymia, depressive symptoms, and psychological distress. LGB minorities were defined as lesbian, gay, bisexual, and men who have sex with men. Exclusion criteria included theses or dissertations, opinion pieces or reviews, and articles that studied use of short message service text messages (not included in our definition of social media). Research in which the sole focus was on gender minority populations (eg, transgender and gender nonbinary) were excluded to avoid conflating results of sexual minorities, which may not be applicable to gender minorities. However, studies in which gender minorities were a subpopulation included in the study LGB sample were included.

Search Process

Literature searches were developed and executed by a health sciences librarian (CBW) in PubMed or MEDLINE (1946-Present), PsycINFO, Ovid (1806-present), and SocINDEX, EBSCOhost (1895-present). Controlled vocabulary from Medical Subject Headings (MeSH), the Thesaurus of Psychological Index Terms, and SocINDEX Subject Terms, along with keywords and descriptors were used for the concepts of lesbian, gay, bisexual, transgender, queer, and interse (LGBTQI), social media, and depression. For all three concepts, we included MeSH and text words in title and abstract. For social media, descriptors included sexual network or partner; seeking sex on internet, online, on the Web, or on websites; finding partners on the internet, online, on the Web, or on websites; and sexual behavior on the internet, online, on the Web, or on websites. For depression, descriptors also included all related depression terms. The Boolean operator "AND" combined the three search components. Searches were limited to journal articles only, with no language or publication year restrictions. The entire list of keywords, descriptors, and search strings used in each database is available as a supplement (Multimedia Appendix 3). Search results were downloaded and imported into an EndNote Library on June 5, 2017. A total of 1259 citations were found. Of these, 539 citations were from PubMed or MEDLINE, 404 citations from PsychINFO, and 316 citations from SocINDEX. There were 160 duplicates records, leaving 1099 citations to screen.

Study Selection and Data Extraction

Screening and data extraction were completed using DistillerSR [49]. Structured forms were uploaded to the software and used throughout the entire process. Six researchers (ALB, CJC, BLH, AS, JES, and DLW) independently screened all article titles and abstracts to generate a set of references for which there was any possibility for selection. Next, these six researchers were divided into three pairs and were randomly assigned an equal number of references; they assessed the full text of these studies to determine eligibility. Interrater reliability was substantial (weighted Cohen kappa, 0.70) [50]. To minimize the risk of reviewer bias, consensus meetings between the first author and each pair of reviewers to resolve differences occurred, but only after independent screening of all articles. In one case, the first author adjudicated a reference for inclusion.

Extraction forms included seven categories of information: (1) study logistics (setting, country, publication year, social media site under study, study design, and funding source); (2) study population characteristics (number of subjects, age, gender, race or ethnicity, sexual minorities included, education level, and income); (3) social media use (number of social networking

sites, time of usage and frequency, scales, and contextual measures); (4) health outcomes measured (primary and secondary outcomes measured and scales); (5) main results and limitations; (6) main tenet of the minority stress theory under study (ie, distal stressors, proximal stressors, and social support); and (7) appropriateness of reporting. To ensure accuracy, we implemented a quality-control mechanism in which one reviewer completed a first data extraction and the second reviewer validated or disagreed with it. Again, disagreements were resolved in consensus meetings with each pair and the first author.

Appropriate Study Reporting

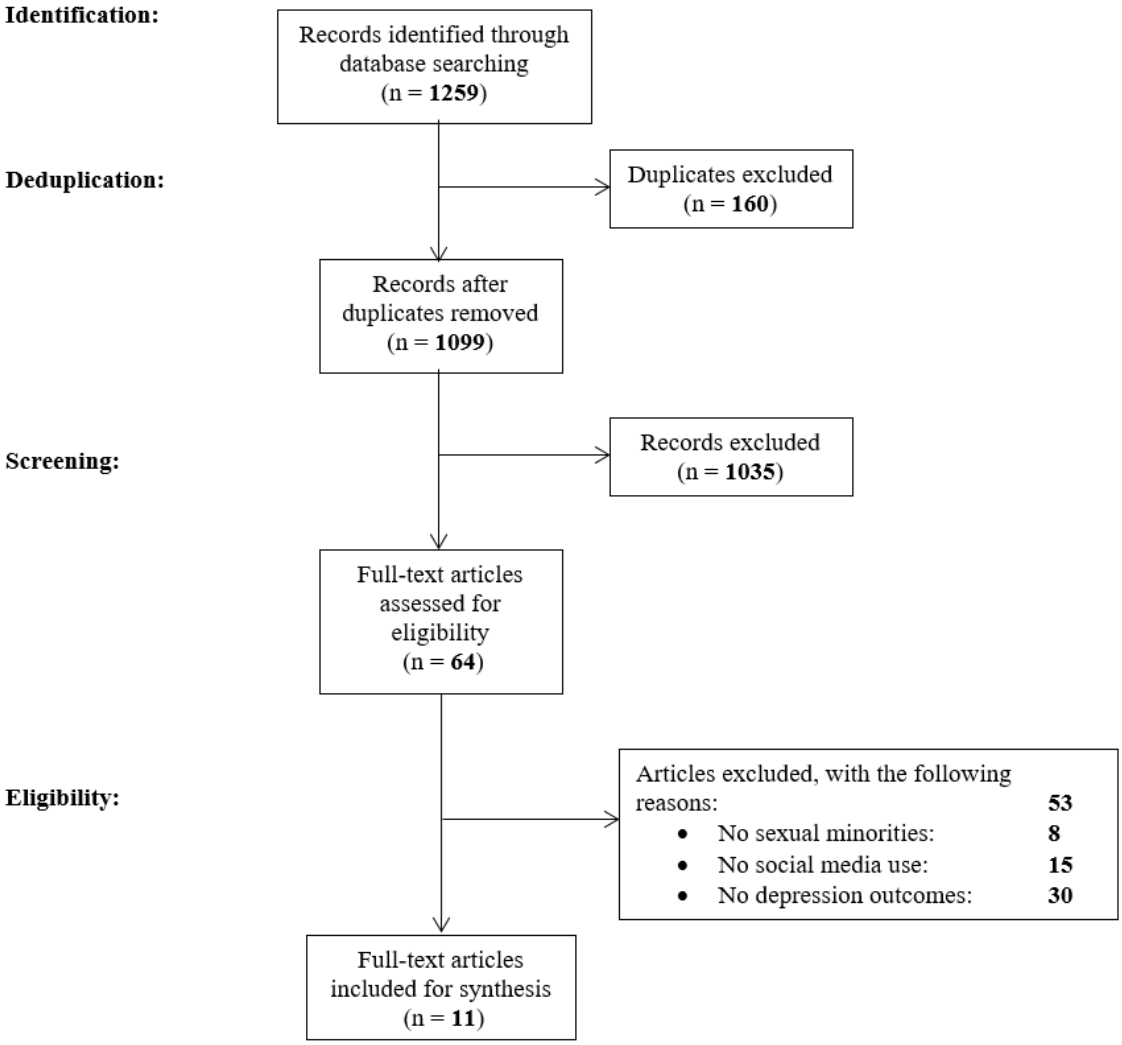
We assessed the appropriate reporting of the included studies. For quantitative studies, we used the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement Checklist v4.0 [51,52]. The STROBE statement consists of a checklist of 22 items related to all sections of research manuscripts; STROBE provides reporting recommendations for studies that investigate associations between exposures and health outcomes [51,52]. We assigned values of 0-1 to each check mark. Thus, total score for each manuscript could range from 0 to 22, in which 22 means the study fully met the STROBE standards of appropriate reporting. For qualitative manuscripts, we used the Consolidated Criteria for Reporting Qualitative Research (COREQ-32) [53], a checklist of 32 items aimed at improving the quality of reporting individual interviews– and focus groups–generated data. We used the same previously explained mechanism to score each manuscript from 0 to 32, in which 32 means the study fully met the COREQ-32 standards of appropriate reporting. Each study was appraised by at least two reviewers, and the first author was consulted to resolve any disagreement. The assessments of appropriate reporting for all manuscripts are available as supplements (Multimedia Appendices 4 and 5).

Results

Study Identification

We identified 1259 records through our database searching process. After excluding duplicates, we reviewed 1099 unduplicated citations (Figure 1).

Figure 1. Flowchart of studies screened and included in a 2017 systematic review of social media use and depression among lesbian, gay, and bisexual minority populations.



Of these, 1035 were excluded after title and abstract screening. Of the 64 full-text manuscripts that were assessed for eligibility, 53 were excluded for different reasons: 8 lacked a focus on sexual minorities, 15 did not specifically assess social media use, and 30 did not have depression as part of the outcomes under study. Eleven research articles were, thus, included in the final sample. Reference lists of included articles were examined for additional studies. However, no new study that met the inclusion criteria was identified using this method.

Study Characteristics and Appropriate Reporting

Of the 11 included studies, 8 (72%) consisted of cross-sectional surveys [54-61], 2 (18%) consisted of qualitative analyses [62,63], and 1 (9%) combined cross-sectional surveys with social network analysis [64] (condensed Table 1; for full table, please see Multimedia Appendix 6). The social media site or platform targeted in the study varied across the included manuscripts. Of the 11 studies, 5 (46%) did not focus on a specific social media site [54,56,57,60,61], 2 (18%) targeted the use of blogs and discussion forums [59,63], and 4 (36%) focused on a specific social media site or platform [55,58,62,64].

Table 1. Characteristics of studies on social media use and depression including sexual minorities published between January 2003 and June 2017.

Author(s), country, year	Design	Social media site or app	Participants			Score[a]
			N	Sample description	Sexual minorities (%)	
Morelli et al, Italy, 2016 [54]	Cross-sectional survey	No specific site	1334	Middle- and high-school students and young adults	Lesbian or gay (12.6)	11[b]
Gibbs & Rice, USA, 2016 [55]	Cross-sectional survey	Grindr	195	Male users of a hook-up mobile app	Gay (86); bisexual (9.8)	20[b]
Cenat et al, Canada, 2015 [56]	Cross-sectional survey	No specific site	6540	Students from 34 participating high schools across Canada	Lesbian or gay (1.3); bisexual (10)	20[b]
Rubin & McClelland, USA, 2015 [62]	Individual interviews	Facebook	8	Female adolescent who reported being daily Facebook users	Lesbian (62.5); bisexual (37.5)	15[c]
Duong & Bradshaw, USA, 2014 [57]	Cross-sectional survey	No specific site	951	Sexual minority students, grades 9-12 from 105 NYC[d] schools	N/A[e]	17[b]
Homan et al, USA, 2014 [64]	Cross-sectional survey; social network analysis	TrevorSpace	195	Users of a LGBQ[f] social networking site	N/A	19[b]
Lester, USA, 2006 [58]	Cross-sectional survey	Bmezine	4700	Users of a body modification website	Lesbian or gay (5); bisexual (37.9)	5[b]
Cooper & Blumenfeld, USA, 2012 [59]	Cross-sectional survey	General blogs and discussion boards	310	National sample of middle- and high-school students who identified as LGB[g], or with same-sex attraction or LGBT[h] allied youth	Lesbian or gay or bisexual (80.6)	13[b]
Alang & Fomotar, USA, 2014 [63]	Netnography	Unidentified forum for new and expecting parents	N/A	Messages from an unidentified online forum for lesbian mothers with postpartum depression	Lesbian (100)	17[c]
Ceglarek & Ward, USA, 2016 [60]	Cross-sectional survey	No specific site	570	College students and community LGBTQ[i] group members	Lesbian or gay (6.8); bisexual (5.4); predominantly heterosexual (13.5)	20[b]
Ramsey et al, USA, 2016 [61]	Cross-sectional survey	No specific site	634	Undergraduate college students	Lesbian or gay (7.6); bisexual (4.4); mostly gay or lesbian (3.5); mostly heterosexual (3.3); questioning (0.8)	19[b]

[a]Appropriate reporting score. Reporting adequacy was assessed using footnotes b and c.

[b]The Strengthening the Reporting of Observational Studies in Epidemiology (range 0-22) for quantitative studies.

[c]The Consolidated Criteria for Reporting Qualitative Research 32 (range 0-32) for qualitative studies.

[d]NYC: New York City.

[e]N/A: not applicable.

[f]LGBQ: lesbian, gay, bisexual, and questioning.

[g]LGB: lesbian, gay, and bisexual.

[h]LGBT: lesbian, gay, bisexual, and transgender.

[i]LGBTQ: lesbian, gay, bisexual, transgender, and questioning.

The studies that captured the age of participants [54-56,59-62,64] reported that ages ranged between 11 and 30 years; 27% (3/11) studies did not report participants' age range [57,58,63]. Furthermore, 64% (7/11) studies included both male and female participants [54,56-61]; in these studies, female participants ranged from 55% to 78%. In addition, 18% studies (2/11) included only female participants [62,63] and 9% (1/11) had an exclusively male sample [55]; 18% (2/11) studies included a sample of transgender participants, with this group comprising 0.9% [59] and 0.8% of participants [61]. Additionally, 18% (2/11) studies included a small sample of gender nonconforming participants. In these studies, the percentage of gender nonconforming participants ranged from 0.7% [58] to 1.6% [61]. Participants' gender was not reported in 9% (1/11) studies [64].

Overall, the reporting of sexual minority participants varied across studies; 46% (5/11) of studies combined gay and lesbian identity [54,56,58,60,61]; 18% (2/11) studies reported a "predominantly or mostly heterosexual identity" category [60,61] and 9% (1/11) combined gay, lesbian, and bisexual identities [59]. Furthermore, 18% (2/11) studies reported that their entire sample comprised LGB participants, but participants' sexual orientation was not broken down into specific categories [57,64].

Finally, the appropriate reporting of results was variable. Among 9 quantitative studies, STROBE scores [51] ranged from 5 to 20 out of 22. Of all, 89% (8/9) studies met the reporting standards on their title, abstract, and introduction sections; 67% (6/9) studies met reporting standards on methods, 22% (2/9) on results, 6 (67%) on discussion, and 2 (22%) on funding source reporting (Multimedia Appendix 4). For 2 qualitative studies, COREQ-32 scores [53] were 15 [62] and 17 [63] out of 32, respectively (Multimedia Appendix 5).

Exposure and Outcome Characteristics and Social Media Sites Studied

Operationalizing of social media use measurement varied across studies, and these findings are summarized in a condensed Table 2 (for full table, please see Multimedia Appendix 7). In 9 quantitative studies, social media use was assessed in variable ways: 33.3% (3/9) studies measured the self-reported experience of cyberbullying [56,57,61], 22.2% (2/9) measured the frequency of social media use [59,60], and 22.2% (2/9) assessed the general use of social media (dichotomously) [55,58]. Furthermore, 44.4% (4/9) studies measured only one of the following: sexting behavior [54], integration of social ties on social media [64], number of social media platforms used [60], and motivation to use social media [60]. Qualitative studies explored the use of Facebook profile management tools [62] and experience of using an online support forum [63].

Similarly, assessment of depression varied across quantitative studies. Depression was operationalized as depressive symptoms in 44.4% (4/9) [55,59-61,64], psychological distress in 22.2% (2/9) [54,56], and suicidality in 33.3% (3/9) [56-58] studies; 22.2% (2/9) studies assessed either engagement in physical fights [57] or emotional responses to cyberbullying (including feelings of depression) [59]. Qualitative studies analyzed depression-related themes including social, emotional, and health consequences of stress caused by managing participants' Facebook profiles [62] and the emotional experience derived from using an online support forum [63].

Main Findings

Exposure to cyberbullying on social media among LGB individuals was frequent, and the majority of those who experienced it reported feelings of depression [59] (condensed Table 2; for full table, please see Multimedia Appendix 7). Compared with heterosexual youth, bisexual boys and girls were more likely to report cyberbullying [56,61]. Among LGB boys and girls, cyberbullying was directly and independently associated with psychological distress [56], depression [61], engaging in physical fights [57], and suicidal thoughts or suicide attempts [56,57]. Compared with heterosexuals, sexual minority users of an online forum group also had higher rates of suicidality [58].

Association between using social media and depression differed depending on which characteristic or pattern of use was under study. One study found that lesbian or gay participants had higher rates of sexting behavior than their heterosexual peers; however, psychological distress was not different across three levels of sexting [54]. Another study found moderate levels of depression among all male users of a gay hook-up mobile app [55]. Yet another study found that social media users who had more friends who knew each other (ie, a tightly integrated social network) predicted lower depression scores than those who did not [64]. When compared with their heterosexual peers, sexual minority youth reported higher rates of both social media sites used and motives to use them [60]. Furthermore, perceived social support on social media among sexual minority youth was negatively associated with loneliness and using social media to discuss LGB issues was negatively associated with anxiety and hostility [60].

Qualitative explorations about social media use and mental health among sexual minorities indicated both risks and benefits. Maintaining a Facebook profile was deemed part of everyday life among lesbian and bisexual females [62]. However, it also requires constant surveillance and monitoring of one's social interactions, which in turn can be a stressor, leading to rumination of ideas, shame, and depression if one is excluded or outed [62]. On the other hand, among LGB mothers dealing with postpartum depression, an online forum served as a space where they could disclose their experiences with the condition while sharing ways to cope with it, building a community that provided different forms of social support [63].

Finally, while most of the reviewed studies assessed only one component of Meyer's theory, other studies focused on more than one. Of the included studies, 81.8% (9/11) focused on social media experiences as a source of stressors, such as victimization or cyberbullying [54-59,61,62,64]; 36.4% (4/11) studies focused on social media as a potential source of support for LGB individuals [55,60,63,64] and 9.1% (1/11) assessed proximal stressors (ie, sexual orientation identity concealment on social media) [62].

Table 2. Exposure and outcome assessment and main findings of studies of social media use and depression including sexual minorities published between January 2003 and June 2017.

Author(s), country, year	Exposure assessment tool	Outcome assessment tool	Aspect of minority stress theory studied
Morelli et al, Italy, 2016 [54]	Modified version of the Sexting Behaviors Scale	12-item General Health Questionnaire	Distal stressors
Gibbs & Rice, USA, 2016 [55]	Sample was recruited exclusively from Grindr (overall use was not assessed)	4-item Center for Epidemiological Studies Depression Scale	Distal stressors; social support
Cenat et al, Canada, 2015 [56]	Item asking, "In the last 12 months, how many times someone has bullied you (rumors, intimidation, threatening, etc) using the internet (Facebook, MySpace, MSN, email, text, etc)?"	10-item Kessler Psychological Distress Scale; item asking, "Have you ever seriously thought of committing suicide?"	Distal stressors
Rubin & McClelland, USA, 2015 [62]	Experience of being young, queer, and a person of color in an online network	Consequences of social exclusionary practices within an online network	Distal stressors; proximal stressors
Duong & Bradshaw, USA, 2014 [57]	Item from Youth Risk Behavior Survey asking, "During the past 12 months, have you ever been electronically bullied, such as through email, chat rooms, instant messaging, websites, or text messaging?"; item asking, "In the past 12 months, have you ever been bullied on school property?"	Item asking, "During the past 12 months, how many times did you actually attempt suicide?"; "During the past 12 months, how many times were you in a physical fight?"	Distal stressors
Homan et al, USA, 2014 [64]	Social network structure graph	9-item Patient Health care Questionnaire	Distal stressors; social support
Lester, USA, 2006 [58]	Sample was recruited exclusively from Bmezine (overall use was not assessed)	Item asking, "How many times have you attempted suicide?"	Distal stressors
Cooper & Blumenfeld, USA, 2012 [59]	Item asking, "How often in an average week do you use communication technologies (eg, blogging, chat rooms, and discussion boards)?"; item asking, "How often in the last 30 days have you been harassed based on your sexual identity?"	Not provided	Distal stressors
Alang & Fomotar, USA, 2014 [63]	Assessment of the role of an online forum as source of social support	Experience of lesbian mothers with postpartum depression using a dedicated online forum	Social support
Ceglarek & Ward, USA, 2016 [60]	Item asking, "How often do you use social networking sites?"; "Which social networking sites do you use?"; "How much these statements apply to you?" Example statement: "I use social networking sites to seek groups of people similar to myself"	26-item Brief Symptom Inventory	Social support
Ramsey et al, USA, 2016 [61]	Cyberbullying; Victimization Scale of the Cyberbullying and Online Aggression Survey	Center for Epidemiological Studies Depression Scale-Revised	Distal stressors

Discussion

Principal Results

In this systematic review, we found a low number of peer-reviewed published research examining social media use and depression among LGB persons. We found ample variation in measurement of social media use and operationalization of these measures. Variability across studies was also found in the definition of sexual minorities as well as conflating sexual and gender minorities in the same study population. The implications of these findings and suggestions for future research are discussed below.

Despite our comprehensive inclusion criteria and systematic online search approach (eg, we included articles that measured depressive symptoms using a psychologic distress scale), there were few studies that examined the relationship between social media use and depression among LGB individuals; 9 studies were cross-sectional and only 2 examined qualitative data. Appropriate reporting of results was variable across the included studies. For quantitative studies, most of the variability was due to incomplete reporting of study results, such as demographics, clinical and social characteristics of participants and reasons for nonparticipation, incomplete report of estimates, and nonreporting of ad-hoc analyses (eg, interactions, sensitivity analysis). On the other hand, most of the variation across the 2 qualitative studies was due to inadequate reporting of sample size, nonparticipant characteristics, sample description, development of interview guides, or data saturation.

We found variation in the included studies' approach to whether assess social media use on a platform-specific approach or for social media as a whole. While social media sites share commonalities, in many aspects they are also very different. For example, certain social media sites are more popular among certain groups than others [65]. Moreover, while some actions and modes of interaction (eg, posting a picture; live streaming;

and commenting on someone's tweet, post, or status update) are actions one can perform across several social media sites, the length of time the picture or video is available, the allowed length of response to a comment, and the audience for these can be very different from one social media site to another. It is also important to consider the motivation to use a given site.

A key finding from this review, adding to other research on the same topic [36,66], is that seeking social support and connectedness might be a potentially strong motivation to use social media among LGB individuals [60,62,63]. Social support is a known protective factor against depression [67-69]. This review supports the need for future research that focuses on assessing the role of variables that describe the quality of the social media experience (eg, active and passive use, motivation to use) in order to understand the effect of social media use on depression among LGB persons.

There was considerable variation in how social media use was operationalized and measured among the included studies, which speaks to the complexity of using social media. In some studies, use was operationalized in terms of frequency, the number of platforms used, and for how long the individuals used them. In other studies, use was measured in terms of characteristics, such as experiences with cyberbullying and use of social media to find camaraderie online. These findings are consistent with those of other systematic reviews linking social media use to mental health outcomes within the general population [36,66]. For example, while some studies have found frequency or volume of social media use to be associated with depression [11,70], these studies do not take into consideration the specific activities undertaken on social media (eg, engaging in contentious interactions or comparing one's self to others) that could be associated with depression. It is possible that behaviors such as scrolling through newsfeeds with little interaction could also be a problematic behavior. Any of these actions—which vary greatly but may all yield differing levels of importance to mental health outcomes—may be categorized as social media use. While studies using various measures of social media use add to the richness and understanding of it, they may lead to false comparisons and mixed results. It may be valuable for future research to conduct scale development studies that focus on social media use as a construct. Additionally, use of clear and transparent language that more accurately defines the measurements of use may be beneficial.

The results of this review echo a body of research that found an elevated prevalence of depression and psychological distress among LGB individuals compared with that among their heterosexual counterparts [71-75]. Our findings point to the variability in experiencing depression and psychological distress in association with social media use for bisexually identified individuals. These findings might be explained by the minority stress theory [38], which posits that individuals with marginalized identities experience stress from their social environment due to social status [76]. The higher rates of depression and psychological distress among LGB persons may be attributed in part to experiencing discrimination, harassment, and victimization because of their sexual orientation. The findings that using social media may be a protective factor articulates the argument of the sense of an LGB community, which suggests that belonging to a larger community may buffer the effects of marginalization [77,78]. In terms of social media, being connected to other LGB individuals may reduce the psychological effects of discrimination, harassment, and victimization these persons experience in the social environment. Nevertheless, the small samples of LGB individuals in these studies limit the ability to determine if any subgroup differences exist in the protective nature of community connectedness.

This review found a strong focus within the literature on environmental and societal stressors that, via social media, may impact depression outcomes among LGB persons. Much less emphasis has been put on understanding the role of online social support and proximal stress processes (eg, expectation of rejection, concealment, and internalized homophobia) in the association between social media use and depression in this group. In addition, it is not clear to what extent experiences lived in the offline world translate to the social media world for LGB persons. For example, while social support is an important protective factor for depression, the findings regarding online social support in the general population have been mixed, with one study reporting lower levels of protection against adverse mental health outcomes [67]and another reporting an improved quality of life and well-being among those who feel socially excluded and seek online social support [79]. Future studies should keep expanding research on the different components of the minority stress theory as they relate to social media use and depression.

Importantly, we found methodological concerns across the included studies regarding the definition of sexual minority individuals, as well as the conflation of results from men and women in the sample, making it difficult to interpret as to which group the said results would apply. We found a lack of clearly defined LGB samples in this area of research. Of the 11 included studies, 2 did not report participants' sexual orientation. Among studies that reported sexual orientation, several grouped gay men and lesbian women into one group. This reduction limits the ability to understand how using social media may be associated with psychological distress and depression in each group separately. Upon further analysis, studies that included both sexual orientation and gender identity often conflated these two groups, making it hard to determine the differences in experiences of depression based on sexual orientation or gender identity. The collapsing of groups complicated our ability to understand the nuanced differences experienced by individuals based on sexual orientation and gender identity [80]. These findings suggest the need for research that includes larger samples of LBG participants to allow the study of lesbian women, gay men, bisexual women, and bisexual men separately, as well as samples that allow distinguishing sexual orientation from gender identity when reporting results. The lack of representativeness extends to other subpopulation differences. For example, half of the studies did not report participant race or ethnicity, and among those that did, the racial composition of the samples was predominantly white. However, research suggests that LGB racial or ethnic minorities have different experiences with both social media use and mental health compared with their white counterparts [65,81,82]. Thus,

considerations should be made to ensure these samples include adequate percentages of LGB racial or ethnic minorities.

None of the studies included in this review had participants older than 30 years of age. While young adults are the group with highest levels of social media use, around 70% of adults aged 30-64 years and over 35% of those aged 65 years and above have at least one social media account [65]. Given the higher risk of depression among the LGB population and the potential dual role of social media, the lack of data from older individuals is concerning. Usage of, interaction with, and experiences concerning social media may be different by age group, and these variations could have differential effects on mental health outcomes. Future research focused on improving the sampling of sexual minority populations should also consider improving sampling across different age groups.

Implications and Future Directions

Internal and social stressors related to minority status are at the core of the minority stress theory [38]. Given the global spread of social media as both a tool and environment within which social interactions occur, we may need a potential expansion of Meyer's theory, one that accounts for the social media experience. Elements of this theory can be applied to LGB-related experiences in the social media environment. However, Meyer's theory was published in 2003, at a time when many of the modern social media sites that are used today did not exist. For example, MySpace started in the same year that Meyer's work was published, and Facebook started the following year; since then, there has been a proliferation of various social media sites and platforms, which have dramatically changed the social interaction landscape of LGB individuals. As social media research progresses, we need to empirically test the relationship between the different components of Meyer's theory, social media use, and depression. This will inform whether social media aggravates or alleviates minority stress and depression, as well as how and to what extent. Moreover, we expect that as the field moves forward, this research could potentially reveal new or modified risk and protective factors for LGB individuals' mental health in ways the minority stress theory could not anticipate. Understanding how virtual and nonvirtual social platforms influence mental health, both independently of and interacting with each other, will be critical to gain a full understanding of the role of the social environment on LGB mental health disparities.

Conclusions

There is a growing concern about the impact of social media use on mental health outcomes. LGB individuals are a well-suited population to study the nature of the relationship between social media use and depression due to the disproportionately high prevalence of both in this group. This systematic review supports the need for research that addresses the role of using social media in the pathway of depression and other mental health outcomes among sexual minority populations. Our findings suggest that social media use may be both a protective and risk factor for these outcomes among LGB individuals. Connections and support gained via social media may buffer the impact of geographic isolation, discrimination, and loneliness that some LGB persons experience in their daily lives. However, the pressure of maintaining a desirable social media presence, negative experiences on social media such as cyberbullying, and certain patterns of use may associate with increased depressive symptoms in this population. Our findings also indicate the need for future research in this field to recruit larger samples, have more consistent definitions of the study populations, better define the social media use construct, and incorporate the social media experience into the conceptualization of psychosocial factors that impact sexual minorities' mental health disparities.

Acknowledgments

We gratefully acknowledge funding from The Fine Foundation (grant number 171R32SM-2018) for the completion of this research; the foundation played no role in the study design; collection, analysis, and interpretation of data; writing of the report; or decision to submit the article for publication.

Authors' Contributions

CGEV conceptualized the review; developed inclusion and exclusion criteria; guided the screening, eligibility, and data extraction process; summarized the data; and wrote the first manuscript draft. DLW, AS, JES, and BLH contributed to the screening, eligibility, and data extraction process; figures and tables; and the discussion section of the manuscript. ALB and CJC contributed to the screening, eligibility, and data extraction process and constructed the tables included in Multimedia Appendices 1, 3, 4, and 5. CBW conducted the online search and contributed to the search methods section. MPM and BAP contributed in developing inclusion and exclusion criteria and contributed to all the sections of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.

[PDF File (Adobe PDF File), 44KB - mental_v5i3e10496_app1.pdf]

Multimedia Appendix 2

PROSPERO research protocol.

[PDF File (Adobe PDF File), 84KB - mental_v5i3e10496_app2.pdf]

Multimedia Appendix 3

Online search strategy.

[PDF File (Adobe PDF File), 65KB - mental_v5i3e10496_app3.pdf]

Multimedia Appendix 4

Evaluation of quantitative studies included using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) criteria.

[PDF File (Adobe PDF File), 34KB - mental_v5i3e10496_app4.pdf]

Multimedia Appendix 5

Evaluation of qualitative studies included using the Consolidated Criteria for Reporting Qualitative Research (COREQ-32).

[PDF File (Adobe PDF File), 28KB - mental_v5i3e10496_app5.pdf]

Multimedia Appendix 6

Characteristics of studies of social media use and depression including sexual minorities published between January 2003 and June 2017.

[PDF File (Adobe PDF File), 35KB - mental_v5i3e10496_app6.pdf]

Multimedia Appendix 7

Exposure and outcome assessment and main findings of studies of social media use and depression including sexual minorities published between January 2003 and June 2017.

[PDF File (Adobe PDF File), 41KB - mental_v5i3e10496_app7.pdf]

References

1. Medley G, Lipari RN, Bose J, Cribb DS, Kroutil LA, McHenry G. 2017. Sexual Orientation and Estimates of Adult Substance Use and Mental Health: Results from the 2015 National Survey on Drug Use and Health Internet URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015.htm [accessed 2018-07-05] [WebCite Cache ID 70gVuBPND]
2. King M, Semlyen J, Tai SS, Killaspy H, Osborn D, Popelyuk D, et al. A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. BMC Psychiatry 2008 Aug 18;8:70 [FREE Full text] [doi: 10.1186/1471-244X-8-70] [Medline: 18706118]
3. la RC, Kretschmer T, Dijkstra JK, Veenstra R, Oldehinkel AJ. Disparities in Depressive Symptoms Between Heterosexual and Lesbian, Gay, and Bisexual Youth in a Dutch Cohort: The TRAILS Study. J Youth Adolesc 2016 Mar;45(3):440-456 [FREE Full text] [doi: 10.1007/s10964-015-0403-0] [Medline: 26748920]
4. Björkenstam C, Björkenstam E, Andersson G, Cochran S, Kosidou K. Anxiety and Depression Among Sexual Minority Women and Men in Sweden: Is the Risk Equally Spread Within the Sexual Minority Population? J Sex Med 2017 Mar;14(3):396-403. [doi: 10.1016/j.jsxm.2017.01.012] [Medline: 28202321]
5. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using Social Media as a Research Recruitment Tool: Ethical Issues and Recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
6. Greenwood S, Perrin A, Duggan M. Pew Research Center. 2016 Nov 11. Social Media Update 2016 URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [accessed 2018-06-06] [WebCite Cache ID 6zyXrb4TW]
7. Duggan M, Ellison NB, Lampe C, Lenhart A, Madden M. Pew Research Center. 2015 Jan 09. Frequency of Social Media Use URL: http://www.pewinternet.org/2015/01/09/frequency-of-social-media-use-2/ [accessed 2018-06-06] [WebCite Cache ID 6zyY7qYHt]
8. Pew Research Center. A Survey of LGBT Americans URL: http://www.pewsocialtrends.org/2013/06/13/a-survey-of-lgbt-americans/ [accessed 2018-06-06] [WebCite Cache ID 6zyYSVL2W]

9. Horvath KJ, Danilenko GP, Williams ML, Simoni J, Amico KR, Oakes JM, et al. Technology use and reasons to participate in social networking health websites among people living with HIV in the US. AIDS Behav 2012 May;16(4):900-910 [FREE Full text] [doi: 10.1007/s10461-012-0164-7] [Medline: 22350832]
10. Community Marketing Inc. LGBTQ Research: CMI's 11th LGBTQ Community Survey URL: https://communitymarketinginc.com/lgbt-research/ [accessed 2018-06-06] [WebCite Cache ID 6zyYeUqvB]
11. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among U.S. young adults. Depress Anxiety 2016 Apr;33(4):323-331 [FREE Full text] [doi: 10.1002/da.22466] [Medline: 26783723]
12. Kim J, Lee JR. The Facebook paths to happiness: effects of the number of Facebook friends and self-presentation on subjective well-being. Cyberpsychol Behav Soc Netw 2011 Jun;14(6):359-364. [doi: 10.1089/cyber.2010.0374] [Medline: 21117983]
13. Primack BA, Shensa A, Escobar-Viera CG, Barrett EL, Sidani JE, Colditz JB, et al. Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults. Comput Human Behav 2017 Apr;69:1-9. [doi: 10.1016/j.chb.2016.11.013]
14. Rosen LD, Whaling K, Carrier LM, Cheever NA, Rokkum J. The Media and Technology Usage and Attitudes Scale: An empirical investigation. Comput Human Behav 2013 Nov 1;29(6):2501-2511 [FREE Full text] [Medline: 25722534]
15. Deters FG, Mehl MR. Does Posting Facebook Status Updates Increase or Decrease Loneliness? An Online Social Networking Experiment. Soc Psychol Personal Sci 2013 Sep 1;4(5) [FREE Full text] [doi: 10.1177/1948550612469233] [Medline: 24224070]
16. Shensa A, Escobar-Viera CG, Sidani JE, Bowman ND, Marshal MP, Primack BA. Problematic social media use and depressive symptoms among U.S. young adults: A nationally-representative study. Soc Sci Med 2017 Dec;182:150-157. [doi: 10.1016/j.socscimed.2017.03.061] [Medline: 28446367]
17. Verduyn P, Lee DS, Park J, Shablack H, Orvell A, Bayer J, et al. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. J Exp Psychol Gen 2015 Apr;144(2):480-488. [doi: 10.1037/xge0000057] [Medline: 25706656]
18. Simoncic TE, Kuhlman KR, Vargas I, Houchins S, Lopez-Duran NL. Facebook use and depressive symptomatology: Investigating the role of neuroticism and extraversion in youth. Comput Human Behav 2014 Nov 1;40:1-5 [FREE Full text] [Medline: 25861155]
19. Teppers E, Luyckx K, Klimstra TA, Goossens L. Loneliness and Facebook motives in adolescence: a longitudinal inquiry into directionality of effect. J Adolesc 2014 Jul;37(5):691-699. [doi: 10.1016/j.adolescence.2013.11.003] [Medline: 24321573]
20. Schou AC, Billieux J, Griffiths MD, Kuss DJ, Demetrovics Z, Mazzoni E, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. Psychol Addict Behav 2016 Mar;30(2):252-262. [doi: 10.1037/adb0000160] [Medline: 26999354]
21. Rosenthal SR, Buka SL, Marshall BDL, Carey KB, Clark MA. Negative Experiences on Facebook and Depressive Symptoms Among Young Adults. J Adolesc Health 2016 Nov;59(5):510-516. [doi: 10.1016/j.jadohealth.2016.06.023] [Medline: 27546886]
22. Davila J, Hershenberg R, Feinstein BA, Gorman K, Bhatia V, Starr LR. Frequency and Quality of Social Networking Among Young Adults: Associations With Depressive Symptoms, Rumination, and Corumination. Psychol Pop Media Cult 2012 Apr 1;1(2):72-86 [FREE Full text] [doi: 10.1037/a0027512] [Medline: 24490122]
23. De Choudhury M, De S. Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity Internet. Proc Eight Int AAAI Conf Weblogs Soc Media. Palo Alto, CA: Association for Advancement of Artificial Intelligence; 2014. URL: https://pdfs.semanticscholar.org/2db7/15a479c8961d3020fe906f7bedfa0311b937.pdf [accessed 2018-06-06] [WebCite Cache ID 6zyZJ4kao]
24. Ernala SK, Rizvi AF, Birnbaum ML, Kane JM, De Choudhury M. Linguistic Markers Indicating Therapeutic Outcomes of Social Media Disclosures of Schizophrenia. In: Proc ACM Hum Comput Interact. 2017 Nov Presented at: Conference: Proceedings of the 21st ACM Conference on Computer-Supported Cooperative Work and Social Computing; 2017; New York, NY p. 1-27. [doi: 10.1145/3134678]
25. de Vries DA, Kühne R. Facebook and self-perception: Individual susceptibility to negative social comparison on Facebook. Pers Individ Diff 2015 Nov;86:217-221. [doi: 10.1016/j.paid.2015.05.029]
26. Lup K, Trub L, Rosenthal L. Instagram #instasad?: exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. Cyberpsychol Behav Soc Netw 2015 May;18(5):247-252. [doi: 10.1089/cyber.2014.0560] [Medline: 25965859]
27. Nesi J, Prinstein MJ. Using Social Media for Social Comparison and Feedback-Seeking: Gender and Popularity Moderate Associations with Depressive Symptoms. J Abnorm Child Psychol 2015 Nov;43(8):1427-1438. [doi: 10.1007/s10802-015-0020-0] [Medline: 25899879]
28. Vogel EA, Rose JP, Roberts LR, Eckles K. Social comparison, social media, and self-esteem. Psychol Pop Media Cult 2014;3(4):206-222. [doi: 10.1037/ppm0000047]
29. Feinstein BA, Hershenberg R, Bhatia V, Latack JA, Meuwly N, Davila J. Negative social comparison on Facebook and depressive symptoms: Rumination as a mechanism. Psychol Pop Media Cult 2013;2(3):161-170. [doi: 10.1037/a0033111]

30. Wang JL, Wang HJ, Gaskin J, Hawk S. The Mediating Roles of Upward Social Comparison and Self-esteem and the Moderating Role of Social Comparison Orientation in the Association between Social Networking Site Usage and Subjective Well-Being. Front Psychol 2017 May 11;8:771 [FREE Full text] [doi: 10.3389/fpsyg.2017.00771] [Medline: 28553256]
31. Grieve R, Indian M, Witteveen K, Anne Tolan G, Marrington J. Face-to-face or Facebook: Can social connectedness be derived online? Comput Human Behav 2013 May;29(3):604-609. [doi: 10.1016/j.chb.2012.11.017]
32. Tandoc EC, Ferrucci P, Duffy M. Facebook use, envy, and depression among college students: Is facebooking depressing? Comput Human Behav 2015 Feb;43:139-146. [doi: 10.1016/j.chb.2014.10.053]
33. Charoensukmongkol P. Mindful Facebooking: The moderating role of mindfulness on the relationship between social media use intensity at work and burnout. J Health Psychol 2016 Sep;21(9):1966-1980. [doi: 10.1177/1359105315569096] [Medline: 25680915]
34. Blackwell L, Hardy J, Ammari T, Veinot T, Lampe C, Schoenebeck S. Proc 20 CHI Conf Hum Factors Comput Syst. New York, New York, USA: ACM Press; 2016. LGBT Parents and Social Media Internet URL: https://yardi.people.si.umich.edu/pubs/Schoenebeck_LGBTParents16.pdf [accessed 2018-06-06] [WebCite Cache ID 6zyakmNH0]
35. Haimson O, Brubaker J, Dombrowski L, Hayes G. Proc 18th ACM Conf Comput Support Coop Work Soc Comput. New York, New York, USA: ACM Press; 2015. Disclosure, Stress, and Support During Gender Transition on Facebook Internet URL: http://oliverhaimson.com/PDFs/HaimsonDisclosureStressSupport.pdf [accessed 2018-06-06] [WebCite Cache ID 6zyatpurU]
36. Seabrook EM, Kern ML, Rickard NS. Social Networking Sites, Depression, and Anxiety: A Systematic Review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]
37. Burke M, Kraut RE. The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength. J Comput Mediat Comm 2016 Jul 26;21(4):265-281. [doi: 10.1111/jcc4.12162]
38. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
39. Marshal MP, Dermody SS, Cheong J, Burton CM, Friedman MS, Aranda F, et al. Trajectories of depressive symptoms and suicidality among heterosexual and sexual minority youth. J Youth Adolesc 2013 Aug;42(8):1243-1256 [FREE Full text] [doi: 10.1007/s10964-013-9970-0] [Medline: 23784511]
40. Rendina HJ, Gamarel KE, Pachankis JE, Ventuneac A, Grov C, Parsons JT. Extending the Minority Stress Model to Incorporate HIV-Positive Gay and Bisexual Men's Experiences: a Longitudinal Examination of Mental Health and Sexual Risk Behavior. Ann Behav Med 2017 Apr;51(2):147-158 [FREE Full text] [doi: 10.1007/s12160-016-9822-8] [Medline: 27502073]
41. Primack BA, Escobar-Viera CG. Social Media as It Interfaces with Psychosocial Development and Mental Illness in Transitional Age Youth. Child Adolesc Psychiatr Clin N Am 2017 Apr;26(2):217-233. [doi: 10.1016/j.chc.2016.12.007] [Medline: 28314452]
42. Jelenchick LA, Eickhoff JC, Moreno MA. "Facebook depression?" social networking site use and depression in older adolescents. J Adolesc Health 2013 Jan;52(1):128-130. [doi: 10.1016/j.jadohealth.2012.05.008] [Medline: 23260846]
43. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011 Jun;28(6):447-455 [FREE Full text] [doi: 10.1002/da.20805] [Medline: 21400639]
44. Torous J, Keshavan M. The role of social media in schizophrenia: evaluating risks, benefits, and potential. Curr Opin Psychiatry 2016 May;29(3):190-195. [doi: 10.1097/YCO.0000000000000246] [Medline: 26967314]
45. Pavalanathan U, De Choudhury M. Identity Management and Mental Health Discourse in Social Media. Proc Int World Wide Web Conf 2015 May;2015(Companion):315-321 [FREE Full text] [Medline: 27376158]
46. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
47. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [Medline: 19622552]
48. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
49. Sykes S. Evidence Partners. 2016 Jun 09. Distiller SR URL: https://blog.evidencepartners.com/citing-distiller-in-publications-and-presentations [accessed 2018-06-06] [WebCite Cache ID 6zybGMtbl]
50. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
51. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577. [Medline: 17938396]

52. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
53. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
54. Morelli M, Bianchi D, Baiocco R, Pezzuti L, Chirumbolo A. Sexting, psychological distress and dating violence among adolescents and young adults. Psicothema 2016 May;28(2):137-142. [doi: 10.7334/psicothema2015.193] [Medline: 27112809]
55. Gibbs JJ, Rice E. The Social Context of Depression Symptomology in Sexual Minority Male Youth: Determinants of Depression in a Sample of Grindr Users. J Homosex 2016;63(2):278-299. [doi: 10.1080/00918369.2015.1083773] [Medline: 26295497]
56. Cénat JM, Blais M, Hébert M, Lavoie F, Guerrier M. Correlates of bullying in Quebec high school students: The vulnerability of sexual-minority youth. J Affect Disord 2015 Sep 01;183:315-321 [FREE Full text] [doi: 10.1016/j.jad.2015.05.011] [Medline: 26047959]
57. Duong J, Bradshaw C. Associations between bullying and engaging in aggressive and suicidal behaviors among sexual minority youth: the moderating role of connectedness. J Sch Health 2014 Oct;84(10):636-645. [doi: 10.1111/josh.12196] [Medline: 25154527]
58. Lester D. Sexual orientation and suicidal behavior. Psychol Rep 2006 Dec;99(3):923-924. [doi: 10.2466/PR0.99.3.923-924] [Medline: 17305212]
59. Cooper RM, Blumenfeld WJ. Responses to Cyberbullying: A Descriptive Analysis of the Frequency of and Impact on LGBT and Allied Youth. Journal of LGBT Youth 2012 Apr;9(2):153-177. [doi: 10.1080/19361653.2011.649616]
60. Ceglarek PJ, Ward LM. A tool for help or harm? How associations between social networking use, social support, and mental health differ for sexual minority and heterosexual youth. Comput Human Behav 2016 Dec;65:201-209. [doi: 10.1016/j.chb.2016.07.051]
61. Ramsey J, DiLalla L, McCrary M. Cyber victimization and depressive symptoms in sexual minority college students. In: J Sch Violence. London, United Kingdom: Taylor & Francis United Kingdom; 2016:483.
62. Rubin JD, McClelland SI. 'Even though it's a small checkbox, it's a big deal': stresses and strains of managing sexual identity(s) on Facebook. Cult Health Sex 2015;17(4):512-526. [doi: 10.1080/13691058.2014.994229] [Medline: 25587894]
63. Alang SM, Fomotar M. Postpartum Depression in an Online Community of Lesbian Mothers: Implications for Clinical Practice. Journal of Gay & Lesbian Mental Health 2015 Jan 23;19(1):21-39. [doi: 10.1080/19359705.2014.910853]
64. Homan CM, Lu N, Tu X, Lytle MC, Silenzio VMB. Social Structure and Depression in TrevorSpace. CSCW Conf Comput Support Coop Work 2014 Dec;2014:615-625 [FREE Full text] [doi: 10.1145/2531602.2531704] [Medline: 28492067]
65. Pew Research Center. 2018 Feb 05. Social Media Fact Sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2018-06-06] [WebCite Cache ID 6zybqKTPX]
66. Baker DA, Algorta GP. The Relationship Between Online Social Networking and Depression: A Systematic Review of Quantitative Studies. Cyberpsychol Behav Soc Netw 2016 Nov;19(11):638-648. [doi: 10.1089/cyber.2016.0206] [Medline: 27732062]
67. Lewandowski J, Rosenberg B, Parks M, Siegel J. The effect of informal social support: Face-to-face versus computer-mediated communication. Comput Human Behav 2011;27(5):1806-1814. [doi: 10.1016/j.chb.2011.03.008]
68. Ybarra ML, Mitchell KJ, Palmer NA, Reisner SL. Online social support as a buffer against online and offline peer and sexual victimization among U.S. LGBT and non-LGBT youth. Child Abuse Negl 2015 Jan;39:123-136. [doi: 10.1016/j.chiabu.2014.08.006] [Medline: 25192961]
69. McConnell EA, Birkett MA, Mustanski B. Typologies of Social Support and Associations with Mental Health Outcomes Among LGBT Youth. LGBT Health 2015 Mar;2(1):55-61 [FREE Full text] [doi: 10.1089/lgbt.2014.0051] [Medline: 26790019]
70. Naja WJ, Kansoun AH, Haddad RS. Prevalence of Depression in Medical Students at the Lebanese University and Exploring its Correlation With Facebook Relevance: A Questionnaire Study. JMIR Res Protoc 2016 May 31;5(2):e96 [FREE Full text] [doi: 10.2196/resprot.4551] [Medline: 27246394]
71. Szymanski D. Heterosexism and Sexism as Correlates of Psychological Distress in Lesbians. J Couns Dev 2005;83(3):360. [doi: 10.1002/j.1556-6678.2005.tb00355.x]
72. Kelleher C. Minority stress and health: Implications for lesbian, gay, bisexual, transgender, and questioning (LGBTQ) young people. Couns Pyschol Q 2009:373-379. [doi: 10.1080/09515070903334995]
73. Russell ST, Fish JN. Mental Health in Lesbian, Gay, Bisexual, and Transgender (LGBT) Youth. Annu Rev Clin Psychol 2016;12:465-487 [FREE Full text] [doi: 10.1146/annurev-clinpsy-021815-093153] [Medline: 26772206]
74. Mustanski B, Andrews R, Herrick A, Stall R, Schnarrs PW. A syndemic of psychosocial health disparities and associations with risk for attempting suicide among young sexual minority men. Am J Public Health 2014 Feb;104(2):287-294. [doi: 10.2105/AJPH.2013.301744] [Medline: 24328641]

75. Marshal MP, Dietz LJ, Friedman MS, Stall R, Smith HA, McGinley J, et al. Suicidality and depression disparities between sexual minority and heterosexual youth: a meta-analytic review. J Adolesc Health 2011 Aug;49(2):115-123 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.02.005] [Medline: 21783042]
76. Kertzner RM, Meyer IH, Frost DM, Stirratt MJ. Social and psychological well-being in lesbians, gay men, and bisexuals: the effects of race, gender, age, and sexual identity. Am J Orthopsychiatry 2009 Oct;79(4):500-510 [FREE Full text] [doi: 10.1037/a0016848] [Medline: 20099941]
77. Doty ND, Willoughby BLB, Lindahl KM, Malik NM. Sexuality related social support among lesbian, gay, and bisexual youth. J Youth Adolesc 2010 Oct;39(10):1134-1147. [doi: 10.1007/s10964-010-9566-x] [Medline: 20593304]
78. Lin YJ, Israel T. Development and validation of a Psychological Sense of LGBT Community Scale. J Community Psychol 2012 Jun 04;40(5):573-587. [doi: 10.1002/jcop.21483]
79. Yao T, Zheng Q, Fan X. The Impact of Online Social Support on Patients' Quality of Life and the Moderating Role of Social Exclusion. J Serv Res 2015 May 08;18(3):369-383. [doi: 10.1177/1094670515583271]
80. Patterson JG, Jabson JM, Bowen DJ. Measuring Sexual and Gender Minority Populations in Health Surveillance. LGBT Health 2017 Apr;4(2):82-105 [FREE Full text] [doi: 10.1089/lgbt.2016.0026] [Medline: 28287877]
81. Mollon L. The forgotten minorities: health disparities of the lesbian, gay, bisexual, and transgendered communities. J Health Care Poor Underserved 2012 Feb;23(1):1-6. [doi: 10.1353/hpu.2012.0009] [Medline: 22643458]
82. Whitfield D, Walls N, Langenderfer-Magruder L, Clark B. Queer Is the New Black? Not So Much: Racial Disparities in Anti-LGBTQ Discrimination. J Gay Lesbian Soc Serv 2014:426-440. [doi: 10.1080/10538720.2014.955556]

Abbreviations

COREQ-32: Consolidated Criteria for Reporting Qualitative Research
LGB: lesbian, gay, and bisexual
LGBQ: lesbian, gay, bisexual, and questioning
LGBT: lesbian, gay, bisexual, and transgender
LGBTQ: lesbian, gay, bisexual, transgender, and questioning
LGBTQI: lesbian, gay, bisexual, transgender, queer, and intersex
MeSH: Medical Subject Headings
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by J Torous; submitted 23.03.18; peer-reviewed by B Bie, SK Ernala, N Machado; comments to author 03.05.18; revised version received 18.05.18; accepted 01.06.18; published 23.07.18.

Please cite as:
Escobar-Viera CG, Whitfield DL, Wessel CB, Shensa A, Sidani JE, Brown AL, Chandler CJ, Hoffman BL, Marshal MP, Primack BA For Better or for Worse? A Systematic Review of the Evidence on Social Media Use and Depression Among Lesbian, Gay, and Bisexual Minorities
JMIR Ment Health 2018;5(3):e10496
URL: http://mental.jmir.org/2018/3/e10496/
doi:10.2196/10496
PMID:30037786

Review

Web-Based Mindfulness Interventions for Mental Health Treatment: Systematic Review and Meta-Analysis

Julia Sevilla-Llewellyn-Jones[1,2,3,4], ClinPsyD, PhD; Olga Santesteban-Echarri[3,5,6], PhD; Ingrid Pryor[3,5], BA (Hons); Patrick McGorry[3,5], PhD; Mario Alvarez-Jimenez[3,5], PhD

[1]Institute of Psychiatry and Mental Health, Health Research Institute (IdISSC). Hospital Clínico San Carlos, Madrid, Spain

[2]Mental Health Department, Clinico Virgen de la Victoria Hospital, Málaga, Spain

[3]Orygen, The National Centre of Excellence in Youth Mental Health, Melbourne, Australia

[4]Faculty of Psychology, Malaga University, Malaga, Spain

[5]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia

[6]Faculty of Health Sciences, Universidad Internacional de la Rioja (UNIR), Madrid, Spain

Corresponding Author:
Julia Sevilla-Llewellyn-Jones, ClinPsyD, PhD
Institute of Psychiatry and Mental Health
Health Research Institute (IdISSC). Hospital Clínico San Carlos
Av/ Prof Martin Lago s/n
Madrid, 28040
Spain
Phone: 34 651563055
Fax: 34 913303574
Email: juliasevill@gmail.com

Abstract

Background: Web-based mindfulness interventions are increasingly delivered through the internet to treat mental health conditions.

Objective: The objective of this study was to determine the effectiveness of web-based mindfulness interventions in clinical mental health populations. Secondary aims were to explore the impact of study variables on the effectiveness of web-based mindfulness interventions.

Methods: We performed a systematic review and meta-analysis of studies investigating the effects of web-based mindfulness interventions on clinical populations.

Results: The search strategy yielded 12 eligible studies. Web-based mindfulness interventions were effective in reducing depression in the total clinical sample (n=656 g=−0.609, *P*=.004) and in the anxiety disorder subgroup (n=313, g=−0.651, *P*<.001), but not in the depression disorder subgroup (n=251, *P*=.18). Similarly, web-based mindfulness interventions significantly reduced anxiety in the total clinical sample (n=756, g=−0.433, *P*=.004) and the anxiety disorder subgroup (n=413, g=−0.719, *P*<.001), but not in the depression disorder group (n=251, g=−0.213, *P*=.28). Finally, web-based mindfulness interventions improved quality of life and functioning in the total sample (n=591, g=0.362, *P*=.02) in the anxiety disorder subgroup (n=370, g=0.550, *P*=.02) and mindfulness skills in the total clinical sample (n=251, g=0.724, *P*<.001).

Conclusions: Results support the effectiveness of web-based mindfulness interventions in reducing depression and anxiety and in enhancing quality of life and mindfulness skills, particularly in those with clinical anxiety. Results should be interpreted with caution given the high heterogeneity of web-based mindfulness interventions and the low number of studies included.

(JMIR Ment Health 2018;5(3):e10278) doi:10.2196/10278

KEYWORDS

mindfulness; anxiety disorder; depressive disorder; internetinternet-based; treatment; meta-analysis; mental health.; systematic review

Introduction

In most countries reporting sufficient data, at some point in their lives, over a third of people meet the criteria for being diagnosed with a mental health disorder [1,2]. The most prevalent psychological disorders are anxiety disorders, followed by mood disorders, externalizing disorders such as attention deficit or hyperactivity disorder or oppositional defiant disorder, and substance use disorders [2]. The average lifetime prevalence for depression and anxiety is 11% and 16%, respectively [2,3], and they tend to co-occur [3]. Furthermore, anxiety and depression disorders have high comorbidity with suicide attempts [4]. Developing and evaluating interventions for these disorders is therefore essential.

Psychological interventions are the treatment of choice for mild to moderate mental health conditions such as depression and anxiety [5-7]. In recent years, mindfulness-based interventions (MBIs) have shown promising results [8,9], and mindfulness-based cognitive therapy (MBCT) is recommended as the treatment of choice for relapse prevention in recurrent depression [10].

Mindfulness derives from Buddhist practice and is described as an intentional and nonjudgmental awareness of the present moment [11]. MBIs are assumed to decrease distress by encouraging individuals to relate to their experiences with acceptance and compassion instead of avoidance, control, or suppression [12,13]. Mindfulness is used in various interventions, each tailored for use with specific populations. Examples include Mindfulness-Based Stress Reduction (MBSR) and MBCT tailored for various target populations. In fact, evidence indicates that MBCT is effective in preventing depressive relapse [14,15]. Other interventions, such as Acceptance and Commitment Therapy (ACT), combine principles of mindfulness and acceptance with treatment components from behavioral therapy and experiential psychotherapy [16].

Over the last few years, interest in MBI efficacy has accrued [17]. From the effectiveness of MBCT in preventing relapse in recurrent depression [18-20] to the application of MBI in other mental health conditions, such as substance use, attention deficit or hyperactivity, and anxiety and depression disorders, research has indicated efficacy [8,9,21,22]. However, two recent meta-analyses have reported conflicting results. Vøllestad, Nielsen, and Nielsen [9] have investigated the effects of MBI on anxiety disorders, reporting a large effect size on reducing anxiety symptoms (g=−0.83) and depressive symptoms (g=−0.72). On the other hand, Strauss, Cavanagh, Oliver, and Pettman [8] investigated the effects of MBI on both anxiety and depressive disorders. In contrast to Vøllestad et al's results, Strauss et al did not find a significant effect of MBI on anxiety disorders (P=.09). However, MBI had a significantly large effect in reducing both depressive symptoms in those with depression (g=−0.73) and in depressive symptoms (g=−0.64) when anxiety and depressive disorders were considered together. Vøllestad et al [9] and Strauss et al's [8] meta-analyses focused on different target populations. Furthermore, Strauss et al [8] considered only studies using a group format, while Vøllestad et al [9] included both individual and group formats. It is therefore important to investigate not only the effects of MBI but also the variables that may influence its effectiveness, including duration of treatment [23], group versus individual format [24,25] and target population [17]. Since these variables were inconsistent between Vøllestad et al [9] and Strauss et al's [8] meta-analyses, the reasons behind their differing findings are difficult to decipher.

The previous decade has witnessed increased use of the internet, which has become more than a simple information and communication tool [26]. With increasing access to novel information and communication technologies in developed countries, a growing number of users resort to the internet for information on, and support for, mental health disorders [27]. This rapid development can be easily understood in the context of the significant advantages of online therapy, such as accessibility, low stigma, and cost effectiveness [28]. Two meta-analyses have shown that psychological interventions delivered via smartphone devices can reduce anxiety and depression symptoms [29,30]. In fact, the National Institute for Health and Care Excellence guidelines consider online cognitive behavioral therapy a first-line treatment for depression and anxiety [31]. Likewise, web-based mindfulness interventions (WMIs) have been developed with promising results [32]. WMIs' potential advantages include reductions both in service costs and demand on mindfulness-trained therapists [33].

WMIs have been designed and applied to healthy participants [34] and individuals with physical illness, such as tinnitus [35], and mental health disorders, such as anxiety or depressive disorders [36,37]. To our knowledge, only one meta-analysis has examined the effectiveness of WMIs in clinical (physical and mental illness) and nonclinical populations [32]. This meta-analysis reported moderate, but significant beneficial impact of WMIs on depression (g=0.29) and anxiety (g=0.22) outcomes. While this meta-analysis had several strengths, such as including only online randomized controlled trials, it combined people with and without mental health disorders, with no separate analysis of WMIs' impact on different populations. The effectiveness of MBI has been shown to differ among target populations [17]. Therefore, it is important to determine the effects of WMIs in clinical contexts and also whether MBI is more effective for various clinical conditions such as anxiety and depression.

The primary aim of this study was to update, systematically compile, and analyze the effectiveness of WMIs in patients with a diagnosed mental health disorder. Secondary aims were to explore whether study variables, including participant characteristics, type of control group, and the design and implementation of the intervention had an impact on the effect of WMIs on this population.

Methods

This review was conducted in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (available upon request) [38].

Data Sources

A systematic search of published studies was performed using the following databases: PubMed, PsycInfo, Web of Science, and Scopus, from inception to March 2018. No restrictions were applied for languages. The abstracts, titles, and keywords of studies were searched using combinations of the following terms: (computer OR cyber OR electronic OR email OR e-mail OR internet OR net OR online OR virtual OR Web OR www OR "social media" OR "social network" OR blog OR forum OR mobile OR smartphone) AND (mindfulness OR self-compassion* OR "compassion-based" OR "acceptance and commitment therapy" OR "acceptance-based" OR "loving kindness" OR "person-based cognitive therapy"). Additional articles were identified by hand searching references of retrieved articles and relevant reviews.

Study Selection

To be included in this meta-analysis, studies must have involved participants with a diagnosed mental health condition using either Diagnostic and Statistical Manual of Mental Disorders [39] or International classification of Diseases [40] criteria.

WMIs were defined as Web-based interventions enabling patient-to-expert communication or internet psychoeducation or therapy. Mobile-based interventions were defined as interventions delivered via mobile phones using short message service (SMS) text messaging, multimedia messaging service, or Web apps. Given the field's early state, we considered WMIs as any intervention that incorporated mindfulness either as a therapy (mindfulness-only therapies) or as a main but not the only component of a therapy (mindfulness integrative therapies). The latter definition includes therapies such as acceptance and commitment therapy [16] or acceptance-based cognitive behavior therapy [41].

Studies investigating traditional face-to-face therapy, delivered via teleconference, mobile phone, audiotape, or CD, or that comprised of only downloading a manual or audio file were excluded [42]. Studies that examined online interventions using mindfulness as a minor component of an eclectic therapy, such as using mindfulness as a relaxation exercise, were not included. We also excluded studies investigating the effects of WMIs on anxiety and depression symptoms in nonclinical samples or in samples with somatic disorders or where diagnosis could not be established. In addition, we did not include studies that used WMIs as an adjunct of face-to-face treatment. Finally, we excluded poster presentations and book chapters.

One reviewer (JSLJ) screened all abstracts to determine initial eligibility. Potentially relevant papers were retrieved for detailed examination. Two reviewers (JSLJ and OSE) then independently assessed the retrieved articles. Any disagreements were resolved through discussion by JSLJ, OSE, and MAJ—all clinical psychologists instructed in mindfulness. If necessary, authors were contacted to determine eligibility against inclusion criteria.

Data Extraction and Analysis

Two reviewers (JSLJ and OSE) independently extracted relevant data from selected studies, as follows:

1. *Characteristics of the study and participants*: author; year of publication; diagnosis and diagnostic criteria; number of participants; gender; mean age; follow-up time in weeks; control group details, including type and duration of treatment for the control group (if applicable).
2. *Characteristics of the intervention*: type of intervention; length of treatment; material used (videos, email, phone, SMS text messaging, and presentations); treatment schedule (daily or weekly); setting (computer- or phone-based); assigned home tasks (if applicable); contact with therapist (if applicable).
3. *Intervention evaluation, dropout rates, and associated variables:* adherence; users' evaluations of usability, attractiveness, and helpfulness of the intervention; dropout rates; variables associated with the use of and engagement with the intervention; adverse events and safety of the intervention.

Assessment of Methodological Quality Procedures

Two reviewers (JSLJ and OSE) independently assessed the methodological quality of each of the studies included. For controlled studies, methodological quality was assessed using the Cochrane Collaboration "risk of bias" tool [43]. For uncontrolled studies, we assessed the following criteria: blinding to study design or purpose and incomplete outcome data [42]. We also assessed the quality of the mindfulness intervention using previously proposed criteria adapted for our purpose [17]. Specifically, we assessed whether the included studies used validated mindfulness or acceptance measures, for example, Five Faces of Mindfulness Questionnaire [44], Acceptance and Action Questionnaire [45], Kentucky Inventory Mindfulness Skills [46], Philadelphia Mindfulness Scale [47]; we also assessed the clinical and mindfulness-specific training of the therapist, if applicable, and of the developer of the online intervention.

Data Analysis

For each comparison between treatment and control groups and for each outcome variable (depression, anxiety, quality of life and functioning, and mindfulness skills), we calculated effect size. When the same outcome was evaluated with multiple scales or domains within the same study, we retained only the most valid measure (see Results section) so that each outcome had one effect size. Effect sizes were pooled for predictors analyzed in ≥3 studies reporting data in a usable format. When not enough data was available, authors were contacted for provision of the necessary additional data.

First, pooled analyses, including all diagnoses, were performed to examine the transdiagnostic impact of WMIs. Only studies with a control group were included in this pooled analysis. For each comparison, Hedge's *g* (a correction of Cohen *d* for small samples) was calculated using means and SDs for each outcome measure. To obtain a pre-post comparison between treatment and control groups, we used the formula $d = (M_1 - M_0) / SD_0$, where M_1 and M_0 are means at post- and pretest, respectively, and SD_0 is the pretest SD. We calculated d for the treatment group (d_T) and the control group (d_C). Each effect size indicates, in SDs, the difference in mean between pre- and posttreatment

for each group. That is, we calculated the main effect size by calculating the difference between d_T and d_C. Effect sizes were estimated using Comprehensive Meta-Analysis software version 2.2 (Biostat, Englewood, NJ, USA) [48]. Finally, in order to interpret data, we followed Cohen's [49,50] recommended benchmarks, wherein an effect size is considered small for a Hedge's g of 0.20, moderate for 0.50, and large for 0.80. We used random-effects models to account for within-study error and variation in true effects across studies [51]. To further assess the robustness of our results, subgroup analyses were performed to examine the differential effects of the type of diagnosis (anxiety and depression). We did not assume a common among-study variance component across subgroups. That is, we did not pool the within-group estimates of tau-squared, as this is the option used by RevMan. Moreover, sensitivity analyses were performed to examine (1) statistical heterogeneity; (2) differences by type of therapy (ie, differences between studies using mindfulness-only therapies vs those using mindfulness integrative therapies); and (3) differences by type of control group, that is, wait list, treatment as usual (TAU), or other active control group. Finally, we analyzed the sustainability of treatment effects over time (ie, differences between comparisons of pretreatment to posttreatment vs pretreatment to follow-up) of those studies with follow-up measures. We did not include control groups in our analysis of sustainability of treatment effects because only one study [37] reported follow-up data for the control group. To test the sustainability of treatment effects, we used the formula $d = (M_1 - M_0) / SD_0$, where for each group, M_1 and M_0 are means at posttreatment and pretreatment, respectively, and SD_0 is the pretreatment SD. That is, we calculated d for pretreatment to posttreatment ($d_{P\text{-}P}$) and for pretreatment to follow-up ($d_{F\text{-}U}$) and calculated the main effect size by calculating the difference between $d_{P\text{-}P}$ and $d_{F\text{-}U}$.

Next, we considered heterogeneity, publication bias, and sensitivity. Heterogeneity was calculated by testing the null hypothesis that the true effect size is the same in all studies using the Q statistic [51]. The I^2statistic explains the percentage of variance in observed effects due to variance in true effects. We assessed variance of true effect sizes using $T^2$and the SD of true effects using T. Publication bias was tested by entering data in a funnel graph (a plot of dispersion between study effect and a measure of study size). A symmetrical inverted distribution of the studies around the mean effect size represented in the funnel would indicate an absence of publication bias [51]. If publication bias exists, it was expected that the smallest studies would report the biggest effect sizes.

Results

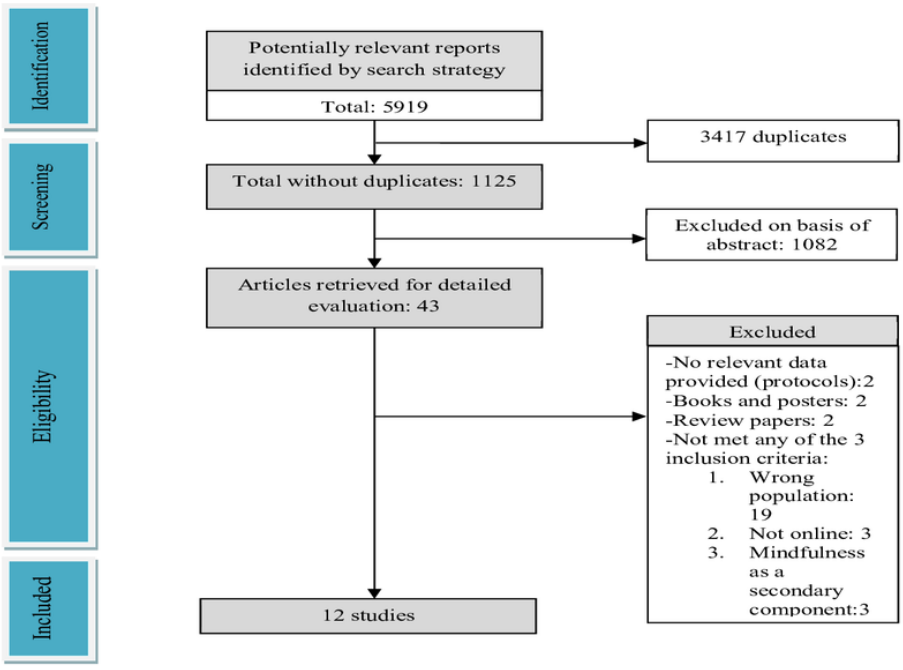
Study Selection

Of 5919 studies retrieved, 12 were included. Figure 1 illustrates study retrieval and selection strategy.

Characteristics of Studies and Participants

A total of 12 studies, involving 919 participants, were selected. Multimedia Appendix 1 depicts the characteristics of these studies, namely, the diagnosis criteria used to determine the eligibility, number of participants, percentage of female participants, mean age, length of treatment, length of follow-up, control group, and length of control. Patients' mean age ranged from 33.2 (SD 10.4) to 46.6 (SD 12.9), and the majority of the sample was female. The main diagnoses were depression and anxiety disorders, frequently diagnosed by a structured interview. All studies reported posttreatment effects, and most studies reported follow-up data, had a control group, and were randomized controlled trials.

Figure 1. Flow chart of study retrieval and selection strategy.



Characteristics of Interventions

Multimedia Appendix 2 depicts the characteristics of the interventions, namely, the type of intervention, length of treatment, materials used, regularity (how often participants were required to log in), setting, the implementation structure (or lack thereof) of the intervention, assigned home task(s), and contact with therapist. Three studies tested mindfulness-only interventions [36,37,52], while four evaluated mindfulness integrative therapies. The duration of treatment varied from 3 to 12 weeks. While the majority of the interventions were flexible, computer-based, followed a modular sequence, and had assigned home tasks, the materials used and the means and frequency of contact with a therapist varied.

Intervention Evaluation, Dropout Rates, and Associated Variables

Table 1 shows the intervention evaluation used in the selected studies. Most studies reported information about adherence and patients' satisfaction (although each used different definitions of these variables). Dropout rates at the end of treatment varied from 0% to 38.5%. Five studies discussed associations between engagement and improved outcomes. Finally, adverse events were rarely reported.

Table 1. Intervention evaluation, dropout rates, and associated variables.

Study	Adherence	Users' evaluation of usability or attractiveness or helpfulness	Dropout rates (%)	Variables associated with increased engagement and better outcomes	Adverse events
Boettcher et al, 2014 [36]	Number of completed mindfulness exercises (homework)	Satisfaction with treatment	11.11	Extensive diagnostic procedure related to adherence and therapeutic change. Clear deadline related to good outcome	Not reported
Ly et al, 2014 [37]	Number of reflections sent to therapist (homework)	Not reported	12.2	Not reported	Not reported
Carlbring et al, 2013 [53]	Number of modules finished and time spent	Not reported	0	Therapist support related to good outcome	Not reported
Kivi et al, 2014 [54]	Number of modules finished and time spent	Not reported	16.67	Extended time to complete module related to lower dropout rates. Therapist support related to better outcomes and lower dropout rates	Not reported
Murray et al, 2015 [55]	Not reported	Qualitative: content, style, negative effects and overall impressions	38.5	Not reported	Body scan meditation generated distress for one participant
Dahlin et al, 2016 [56]	Number of modules finished with homework assignment	Satisfaction with treatment or supportiveness of therapist	19.2	More pictorial information than text and extend time to complete module related to increased engagement	Not reported
Gershkovich et al, 2016 [57]	Completion of modules on weekly basis	Satisfaction with treatment or therapist or ease of use	0	Mail to remember to finish module and postpone video conference	Technical issues with videoconferences
Gershkovich et al, 2017 [58]	Number of modules finished	Satisfaction: treatment or therapist or perceived effectiveness or ease of use	31	Participants with therapist support related to increased engagement	Technical issues in videoconferences
Houghton 2008 [52]	Not reported	Not reported	27.59	Not reported	Not reported
Ivanova et al, 2016 [59]	Number of modules finished	Not reported	Not reported	Extensive use of technology related to low adherence. Personalized feedback increased the adherence to the smartphone platform in comparison to those who did not have personalized feedback	Not reported
Johansson et al, 2013 [60]	Number of modules finished	Satisfaction: amount of text or demand of the treatment or worth the effort	0	Not reported	Not reported
Strandskov et al, 2017 [61]	Not reported	Not reported	21.7	Not reported	Not reported

Table 2. Quality of the interventions.

Study	Use of validated mindfulness or ACT[a] measures	Clinical training of the therapist	Mindfulness training of the therapist	Clinical training of the developer	Mindfulness training of the developer
Boettcher et al, 2014 [36]	No	Not reported	Not reported	General practitioner	Mindfulness-based cognitive therapy, mindfulness-based stress reduction, and other mindfulness training
Ly et al, 2014 [37]	AAQ-II[b]	(4th year) Clinical Psychology MSc; supervised	Not specific—as part of their training	Not applicable—platform already designed	Not reported
Carlbring et al, 2013 [53]	No	Clinical Psychology MSc; supervised	Not specific—as part of their training	Licensed psychologist	Functional contextualism and clinical behavior analysis
Kivi et al, 2014 [54]	No	Licensed psychologist or psychotherapist; supervised	Specific training for the study	Licensed psychologist	Functional contextualism and clinical behavior analysis
Murray et al, 2015 [55]	No	Not applicable	Not reported	Clinicians, costumers, and researches	Mindfulness and ACT
Dahlin et al, 2016 [56]	No	Psychologist graduate students; supervised	Not specific, but some in their training	Clinical psychologist	ACT workshops
Gershkovich et al, 2016 [57]	AAQ-II and Philadelphia Mindfulness Scale	Clinical Psychology doctoral student who received extensive training	ACT intensive training	Clinical Psychology doctoral student	ACT
Gershkovich et al, 2017 [58]	No	Clinical Psychology doctoral student who received extensive training	Not reported	Not reported	Not reported
Houghton 2008 [52]	Kentucky Inventory Mindfulness Skills	Not applicable	Not reported	Not reported	Not reported
Ivanova et al, 2016 [59]	No	Clinical Psychology MSc; supervised	Not reported	Not reported	Not reported
Johansson et al, 2013 [60]	Five Faces of Mindfulness Questionnaire	(3rd year) Clinical Psychology doctoral MSc; supervised by experienced psychotherapist	Clinical training in affect-focused psychodynamic psychotherapy	Clinical psychologist	Affect-focused psychodynamic psychotherapy
Strandskov et al, 2017 [61]	No	(4th year) Clinical Psychology MSc; supervised	Not reported	Not reported	Not reported

[a]ACT: Acceptance and Commitment Therapy.

[b]AAQ-II: Acceptance and Action Questionnaire.

Quality of Interventions

Table 2 depicts the quality of the interventions used in the selected studies. Only four studies used a validated measure of mindfulness skills. All therapists were either graduate psychology students (master's or doctoral level) or licensed clinical psychologists. While the training of therapists on MBI varied, the majority of developers were clinical psychologists with training in MBI.

Methodological Quality

There were 10 randomized controlled trials and two uncontrolled studies included in this meta-analysis. The risk of selection bias was low in all studies, as an online random allocation service independent of the investigators was used. Considering the nature of the interventions, blinding of participants and personnel was not fulfilled by any study. Blinding the outcome assessment criteria was achieved by all studies, except one [54], which was the only study that did not use computer-based assessments. Finally, in both controlled and uncontrolled studies, attrition biases were assessed as low risk.

Meta-Analytic Results

Nine outcomes were identified (stress, health, insomnia, worry, emotional processing, anxiety, depression, quality of life and functioning, and mindfulness skills), with the last four assessed in ≥3 studies.

Web-Based Mindfulness Interventions and Depression Outcome

Depression was examined by 11 studies, of which 5 included participants with a primary diagnosis of an anxiety disorder [36,56-59], 3 with a depressive disorder [37,53,54], 1 with both anxiety and depressive disorders [60], 1 with bipolar disorder [55], and 1 with bulimia disorder [61]. Usable data for meta-analysis of studies with a control group could be retrieved for 8 of these studies [36,37,53,54,56,59-61]. Johansson et al

[60] reported data divided into two subgroups: those with a diagnosis of depression and those with a diagnosis of anxiety. We therefore considered these groups as independent samples. A summary of effect sizes is shown in Figure 2. We found significant large overall effect of WMIs on reducing depression for the pooled sample (n=656, g=−0.609, 95% CI −1.028 to −0.189, P=.004). Significant heterogeneity was noted (Q=55.191, df=8, P<.001, I^2=85.505, T^2=0.348, T=0.590). Subgroup analyses indicated that WMIs had a significant large effect on reducing depression among participants with a diagnosis of anxiety (n=313, g=−0.651, 95% CI −0.945 to −0.356, P<.001), with no evidence of heterogeneity (Q=4.928, df=3, P=.18, I^2=39.119, T^2=0.035, T=0.187). Conversely, the effect of mindfulness treatment on depression was not significant among those with a diagnosis of depression (n=251, g=−0.690, 95% CI −1.694 to −0.313, P=.19). Significant heterogeneity was noted (Q=42.996, df=3, P<.001, I^2=93.023, T^2=0.974, T=0.987). We performed sensitivity analyses to examine the difference between studies including mindfulness-only therapies [36,37] and those including mindfulness integrative therapies [53,54,56,59-61]. Exploratory sensitivity analysis indicated that mindfulness integrative therapies had a significant effect on reducing depression while mindfulness-only therapies did not. Furthermore, we examined differences between studies by type of control group (ie, wait list [56,59-61], TAU [54], or other active control group [36,37,53]). We found a significant effect of WMIs on only reducing depression when compared to wait list. There was not a significant difference when compared to TAU or other active control groups. For studies reporting follow-up data, analysis of the stability of treatment effects indicated that changes were stable over time (analyses are available in Multimedia Appendices 3, 4, and 5, respectively). For all outcomes, exploratory subgroup analysis should be interpreted with caution because having <5 studies per group is likely to provide an imprecise estimation [51].

Web-Based Mindfulness Interventions and Anxiety Outcome

Anxiety was examined in all studies. Of these studies, 7 included participants with a primary diagnosis of an anxiety disorder [36,52,56-60], while others included participants diagnosed with a depressive disorder in addition to anxiety [37,53,54,60], bipolar disorder [55], or bulimia [61]. Usable data of studies with a control group could be retrieved only for 9 of them [36,37,52-54,56,59-61]. Pooled effect sizes are presented in Figure 3. We found a significant moderate effect of WMIs on reducing anxiety for the pooled sample (n=756, g=−0.433, 95% CI −0.725 to −0.141, P=.004). Heterogeneity was noted (Q=35.972, df=9, P<.001, I^2=74.981, T^2=0.165, T=0.406). Subgroup analysis indicated that WMIs had a significant moderate effect on reducing anxiety among participants with a diagnosis of anxiety (n=413, g=−0.719, 95% CI −1.055 to −0.383, P<.001), with evidence of statistical heterogeneity (Q=11.109, df=4, P=.03, I^2=63.994, T^2=0.093, T=0.305). Conversely, the effect of WMIs on anxiety was not significant among those with a diagnosis of depression (n=251, g=−0.213, 95% CI −0.597 to −0.170, P=.28), with no evidence of statistical heterogeneity (Q=7.131, df=3, P=.07, I^2=57.928, T^2=0.088, T=0.297).

Figure 2. Results of depression outcome for the pooled and subgroup samples. MF: mindfulness intervention group.

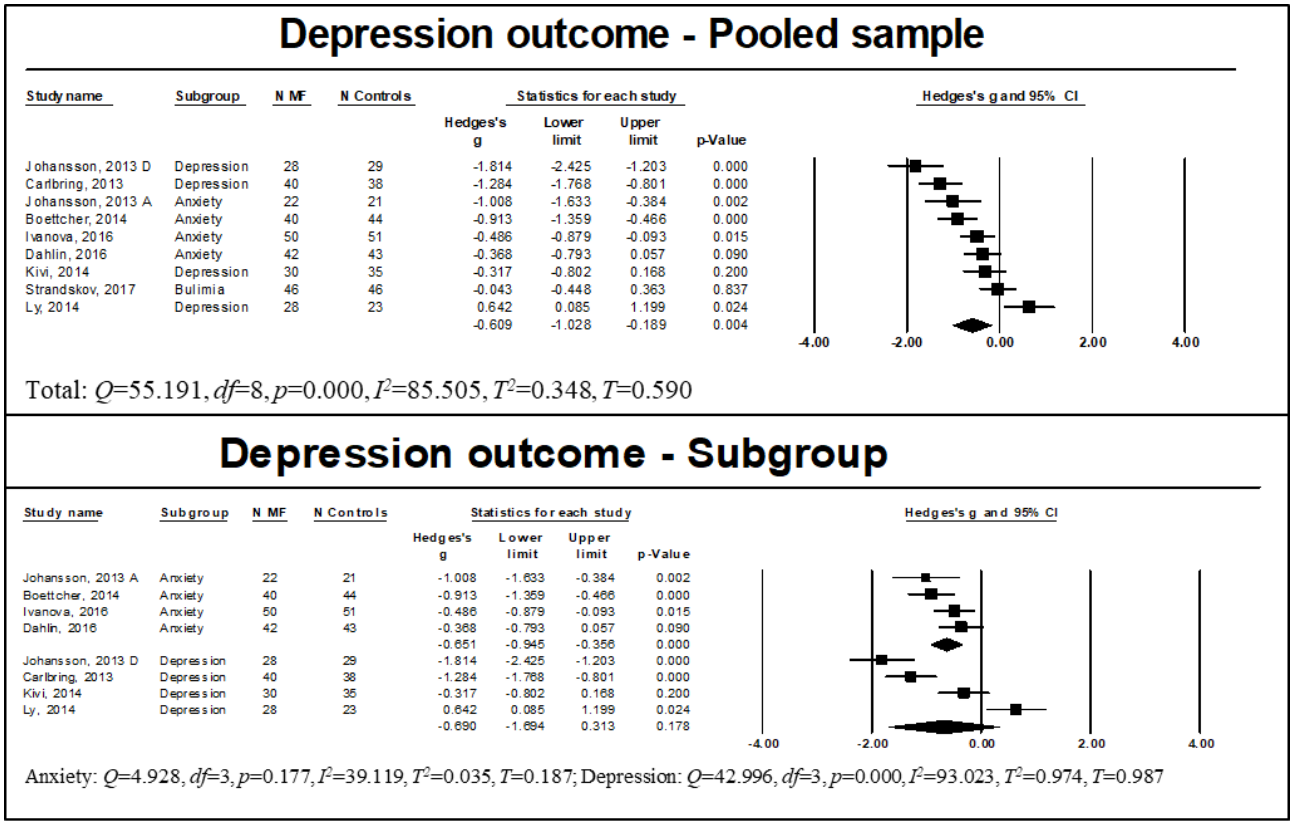
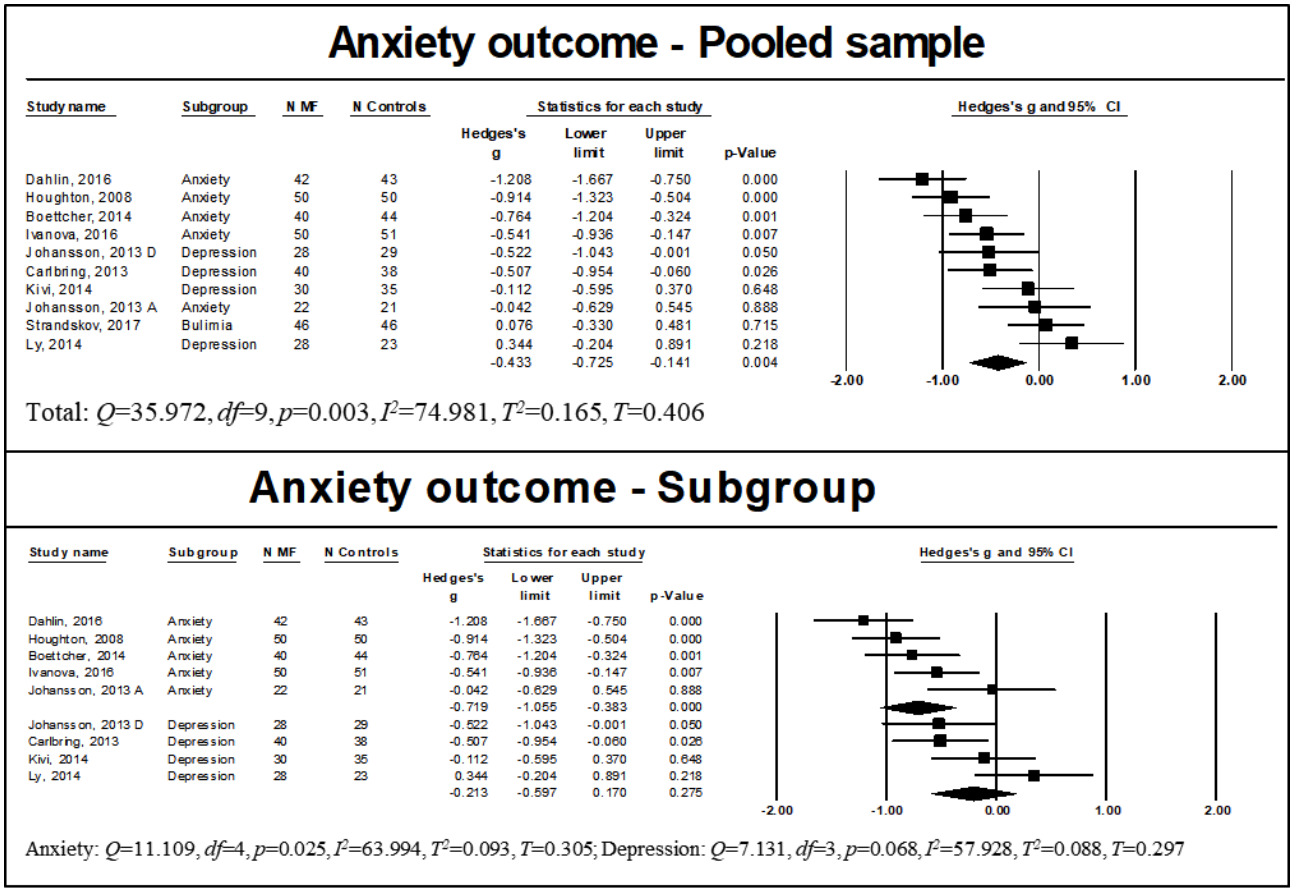


Figure 3. Results of anxiety outcome for the pooled and subgroup samples. MF: mindfulness intervention group.



We performed sensitivity analyses to examine the difference between studies including mindfulness-only therapies [36,37,52] and those including mindfulness integrative therapies [53,54,56,59-61]. Preliminary results indicated that mindfulness integrative therapies had a significant effect on reducing anxiety, while mindfulness-only therapies did not. Mindfulness integrative therapies showed no heterogeneity, while mindfulness-only therapies showed evidence of statistical heterogeneity. Furthermore, we examined differences between studies by type of control group. We found significant effect of WMIs on reducing anxiety when compared to wait list, while there was not a significant difference when compared to TAU or other active control groups. For studies reporting follow-up data, analysis of the stability of treatment effects indicated that changes were stable over time (analyses are available in Multimedia Appendices 6, 7, and 8, respectively).

Web-Based Mindfulness Interventions and Quality of Life and Functioning Outcomes

Quality of life and functioning were examined by 10 studies, of which 4 included participants with anxiety disorders [36,52,56-59,61], 2 included participants with depressive disorders [37,53] and 1 included participants with bipolar disorder [55]. Usable data for meta-analysis of studies with control groups could be retrieved for 7 of them [36,37,52,53,56,59,61]. A summary of effect sizes is showed in Figure 4. We found a significant effect of WMIs on quality of life and functioning for the pooled sample (n=591, g=0.362, 95% CI 0.049 to 0.674, P=.02). Heterogeneity was noted (Q=21.855, df=6, P=.001, I^2=72.546, T^2=0.128, T=0.358). Subgroup analysis indicated a significant, moderate effect of WMIs on quality of life and functioning for those with a diagnosis of anxiety (n=370, g=0.550, 95% CI 0.083 to 1.017, P=.02), and heterogeneity was noted (Q=15.090, df=3, P=.002, I^2=80.120, T^2=0.182, T=0.426). However, there was no significant effect for those with a diagnosis of depression (n=129, g=0.104, 95% CI −0.238 to −0.446, P=.55), with no evidence of significant heterogeneity (Q=0.645, df=1, P=.42, I^2=0.000, T^2=0.000, T=0.000).

We performed subgroup analysis to examine differences between studies including mindfulness-only therapies [36,37,52] and those including mindfulness integrative therapies [53,56,59,61]. Subgroup analysis indicated that neither mindfulness integrative therapies, nor mindfulness-only therapies had a significant effect on quality of life and functioning. Studies investigating mindfulness integrative therapies showed no heterogeneity, while mindfulness-only therapies showed heterogeneity, which could be explained by baseline diagnosis (ie, anxiety vs depression) [36,37,52]. Furthermore, we examined differences between studies by type of control group. We did not find a significant effect of MBI on quality of life and functioning when compared to wait list or other active control groups. For studies reporting follow-up data, analysis of the stability of treatment effects indicated that changes were stable over time (analyses are available in Multimedia Appendices 9, 10, and 11, respectively).

Figure 4. Results of quality of life outcome for the pooled and subgroup samples. MF: mindfulness intervention group.

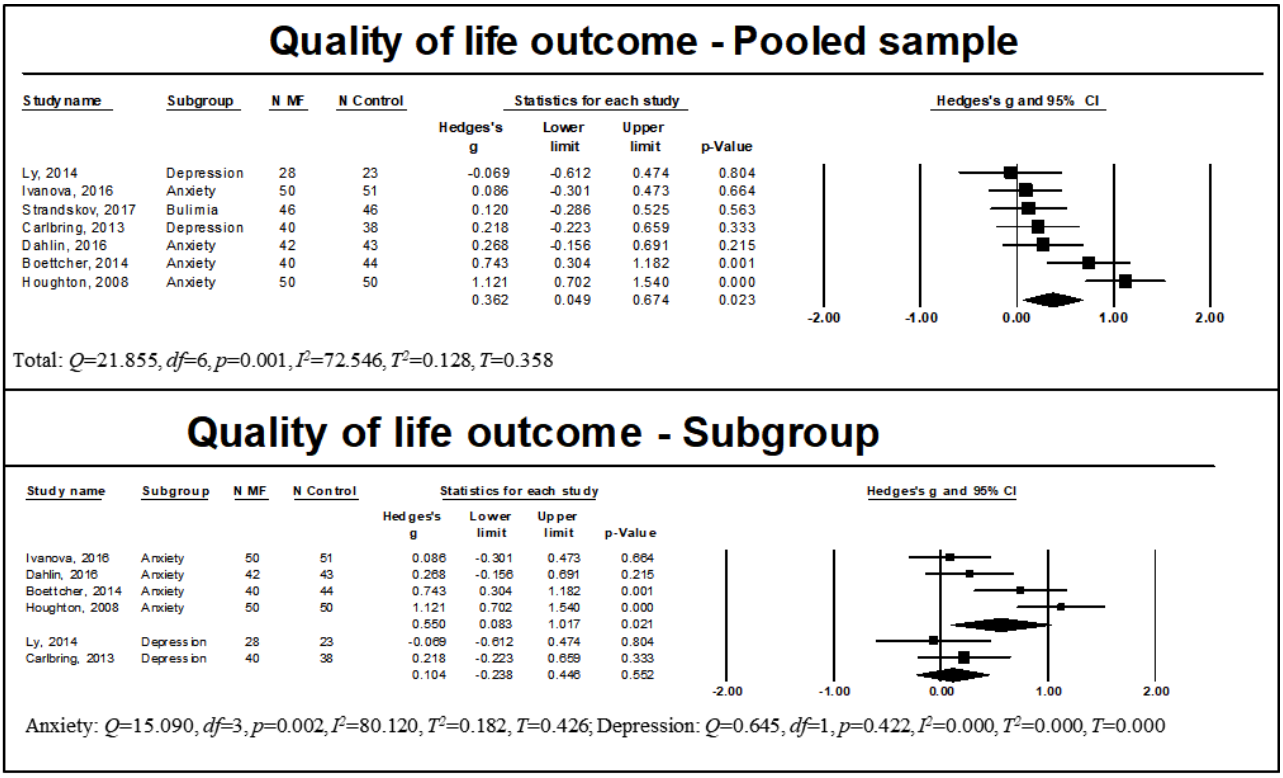
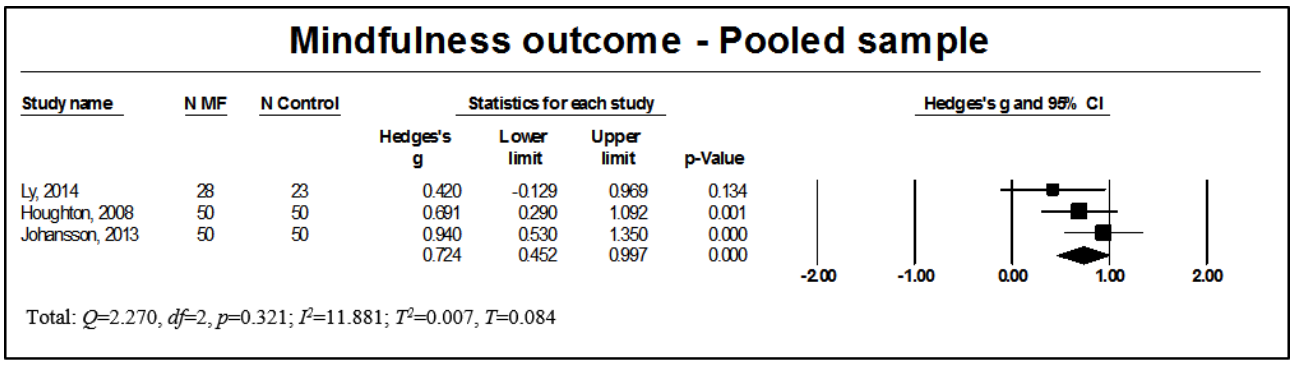


Figure 5. Results of mindfulness outcome for the pooled sample. MF: mindfulness intervention group.



Web-Based Mindfulness Interventions and Mindfulness Skills Outcomes

Mindfulness skills were examined in 4 studies, of which 3 included participants with anxiety disorders [52,57,60] and 2 included participants with depressive disorders [37,60]. Usable data for meta-analysis could be retrieved only for 3 of these [37,52,60]. A summary of effect sizes is shown in Figure 5. We found a significant large effect of WMIs on mindfulness skills for the pooled sample (n=251, g=0.724, 95% CI 0.452 to 0.997, P<.001). No heterogeneity was noted (Q=2.270, df=2, P=.32, I^2=11.881, T^2=0.007, T=0.084). Subgroup analysis could not be calculated due to the small sample size and heterogeneity of participants' diagnoses. For studies reporting follow-up data, analysis of the stability of treatment effects indicated that changes were stable over time (all analyses are available in Multimedia Appendix 12).

Publication Bias

Analysis of the funnel plot indicated some evidence of publication bias for the anxiety outcome. No publication bias was found for depression, quality of life and functioning, or mindfulness skills (Multimedia Appendices 13, 14, 15, and 16, respectively).

Discussion

Overview

The aim of this systematic review and meta-analysis was to examine the clinical and psychosocial effects of WMIs in patients with diagnosed mental health disorders. Overall, our results indicated that WMIs effectively reduced depression and anxiety symptoms and increased quality of life and functioning and mindfulness skills. The secondary aim of this study was to explore factors that can moderate the effects of WMIs in this population. In this respect, preliminary analyses provided initial evidence that WMIs may be particularly beneficial in patients

with anxiety disorders, that mindfulness integrative online therapies may be more effective than mindfulness-only therapies and that WMIs may not be more effective than active control interventions.

Effects of Mindfulness-Based Online Interventions on Depressive Symptoms

In relation to our primary aim, meta-analysis showed that WMIs reduced depressive symptoms in patients diagnosed with mental health conditions. However, further secondary analysis indicated that reduction of depressive symptoms was significant only for patients diagnosed with an anxiety disorder, and not for those with a diagnosis of depression. These findings are consistent with the results of a previous meta-analysis, which found a significant reduction in depressive symptoms associated with face-to-face MBCT in mental health patients with depression, anxiety, and bipolar disorder, when all were studied together [18]. Furthermore, our results are in keeping with Vøllestad et al's [9] meta-analysis of face-to-face mindfulness interventions in patients diagnosed with anxiety, in which the researchers found a large reduction in comorbid depression symptoms associated with the use of mindfulness interventions. One possible explanation for these results could be that reduction of anxiety symptoms in those with a primary diagnosis of anxiety leads to reduced associated depressive symptoms. Alternatively, the low number of studies and statistical heterogeneity might explain the nonsignificant reduction of depressive symptoms in the depression subgroup, especially since the *P* value approaches significance.

Initial research showed that MBCT was effective only in those with >3 prior depressive episodes [19]. However, more recent studies have questioned these findings, indicating that MBCT may also be effective in those with one or two previous depressive episodes [62]. Unfortunately, with the exception of one study [37], we were unable to determine whether the study participants were experiencing their first episode of depression or had experienced recurrent depressive episodes. Thus, we could not perform a subgroup analysis to examine the differential effect of WMIs as a function of number of previous episodes. Future studies should examine the differential effectiveness of WMIs for different stages of illness.

In relation to the secondary analyses undertaken, results indicated that studies that used mindfulness integrative therapies reported significant reductions in depressive symptoms, while those that used mindfulness-only therapies did not. Although this finding indicates that implementing mindfulness techniques alongside other therapies may be more effective in reducing depressive symptoms than mindfulness alone, the low number of studies investigating mindfulness-only therapies requires that we interpret this result with caution.

In relation to the secondary finding that WMIs may have a significant effect in reducing depression when compared to wait list, but not when compared to active control groups, perhaps WMIs are not inferior to other active interventions in depression. Alternatively, this finding may be explained by the significant heterogeneity in the interventions and control groups used. One possible explanation may be that using a wait-list control may have a nocebo effect [63]. That is, using this control condition may result in detrimental effects in the control group and consequently larger effect sizes when compared to the treatment condition. Another possibility is that the digital placebo effect, by which digital interventions have increased effect sizes due to participants' investment in the intervention [64], resulted in inflated effect sizes when an online intervention was compared with an offline control.

Effects of Mindfulness-Based Online Interventions on Anxiety Symptoms

WMIs effectively reduced anxiety symptoms for the overall sample. However, secondary analyses indicated that this reduction was significant only for patients diagnosed with an anxiety disorder, and not for those diagnosed with depression. A recent meta-analysis has reported inconsistent findings on the effects of WMIs on anxiety, with effect sizes varying from small to large [32]. In contrast to our results, Strauss et al [8] found that face-to-face MBI did not reduce anxiety symptom severity among patients with an anxiety disorder. A larger meta-analysis that considered face-to-face mindfulness with patients diagnosed with anxiety disorders did find a large effect size [9]. The inconsistent results of these meta-analyses may be due to differing inclusion criteria, such as having a more inclusive definition of MBI [9] and using only group-based interventions [8], which may influence effects of mindfulness [23].

Secondary analyses indicated that WMIs applied to patients diagnosed with depression did not reduce anxiety symptoms. This may indicate that when anxiety symptoms are secondary to a primary diagnosis (in this case, depression), WMIs are less effective in reducing these symptoms. Alternatively, the low number of studies included could explain the nonsignificant reduction in anxiety symptoms in the depression subgroup results since the *P* value was approaching significance. Additional secondary analyses suggested that mindfulness integrative therapies might work better than mindfulness-only therapies, and WMIs appear to be more effective in the reduction of anxiety symptoms than a wait-list control (but not significantly inferior to other active conditions). As noted above, this may be explained by the heterogeneity in the intervention and control groups or the possibility of a nocebo or digital placebo effect influencing this result [63,64].

Effect of Web-Based Mindfulness Interventions on Quality of Life and Functioning

WMIs significantly improved quality of life and functioning for the overall sample. However, secondary analyses indicated a significant improvement in quality of life and functioning for the anxiety disorder subgroup, but no improvement in the depressive disorder subgroup. Prior research has also shown varying results in relation to this outcome. Spijkerman, Pots, and Bohlmeijer [32] found a small effect size of WMIs on quality of life. In their meta-analysis, Vøllestad et al [9] found a large effect size when investigating the effect of face-to-face MBI on quality of life in patients with an anxiety disorder. Inconsistent findings may be explained by several variables, such as method of program delivery (ie, internet [32] or face-to-face [9]) or differences, such as diagnosis, between the samples of the trials included.

Effect of Web-Based Mindfulness Interventions on Mindfulness Skills

We found a large effect of WMI on increasing mindfulness skills in the overall sample. The improvement in mindfulness skills found in our meta-analysis was consistent with Spijkerman et al's [32] meta-analysis; however, we found a larger effect size. A possible reason is that our meta-analysis focused on a specific population of individuals diagnosed with a mental health condition, while Spijkerman et al's [32] meta-analysis used broader inclusion criteria. In fact, Khoury, Lecomte, Gaudiano, and Paquin [17] have noted the importance of considering sample characteristics since MBIs have been found to be more effective in treating mental health conditions than physical or medical conditions.

Strengths and Limitations

Our study is the first meta-analysis of WMIs focused on patients with diagnosed mental health conditions. We carefully assessed the quality of studies and mindfulness interventions, in line with recommendations by Higgins and Green [43] and Khoury et al [17]. Furthermore, inclusion of secondary analyses allowed for preliminary examination of variables that may impact the effectiveness of WMIs in this population.

A number of methodological issues should be considered for future research. The findings of this meta-analysis are limited by the small sample sizes of the studies included and the heterogeneity of WMIs. Specifically, the mindfulness interventions evaluated in the included studies varied in terms of the regularity of the program, whether homework was provided, and the extent of contact with therapist, among others. Given the low number of studies, we were unable to control for these variables. In addition, results for secondary analyses for anxiety and depression groups should be considered with caution, given the high comorbidity between these disorders [3]. Moreover, significant statistical heterogeneity in the active control condition subgroup limits our exploratory analysis of the relationship between type of control group and WMIs' effectiveness. Notwithstanding these limitations, this meta-analysis provides initial evidence that WMIs can be an effective intervention for those with clinical anxiety and depression, with exploratory analyses indicating important areas for future research into variables that may moderate the impact of WMIs.

Future Research

This meta-analysis revealed marked heterogeneity in the uptake and use of WMIs. This issue is of clinical relevance because research into online interventions has consistently demonstrated high attrition rates [65]. Future research should report what proportion of, and the degree to which, patients engage with different aspects of WMIs over time. Furthermore, noncompletion and good engagement should be measured according to a priori established criteria, and the design, content and interface aspects of WMIs should be carefully analyzed to study their potential differential effects. This will allow for the identification of variables that influence usage rates and treatment effects, and allow clinicians and researchers to tailor implementation of WMIs to maximize engagement and positive outcomes. This will facilitate a fuller understanding of what works well, and for whom. In addition, further studies should investigate the impact of WMIs in different mental health populations, as previous research has indicated the potential benefits of WMIs in young people at ultra-high risk of developing psychosis [66] and face-to-face MBI has been found effective for individuals with psychosis [67].

Currently, with the field of WMIs still in its early stages, the term WMI captures a broad, complex, and poorly defined class of interventions. Studies included in our meta-analysis did not clearly state the focus of each MBI (ie, to reduce stress, prevent depression relapse, or enhance well-being). In keeping with Crane et al's [68] recommendations, future WMIs should follow clearly delineated protocols designed to ensure their quality and integrity. As such, further studies should clearly describe the focus, characteristics and expected therapeutic mechanisms of WMIs. This will not only help to better investigate the efficacy of WMIs (by allowing for comparison of interventions with differing therapeutic goals) but also to examine whether WMIs are exerting their effects via hypothesized treatment mechanisms. Ultimately, this will help inform the adaptation of specific WMIs to meet the needs and preferences of defined clinical populations, as well as to enhance their therapeutic impact by targeting therapeutic mechanisms more specifically.

Finally, future research should determine the effects of theory-driven, targeted WMIs when compared with (1) interventions controlling for unspecific therapeutic factors (eg, attention control) and (2) active interventions targeting differential mechanisms. In order to account for potential nocebo or digital placebo effects, further studies should include either attention control or active control groups. This is further supported by emerging evidence that mindfulness interventions are not inferior to traditional interventions such as cognitive behavioral therapy in the treatment of clinical depression or anxiety [17,69].

Conclusion

In conclusion, our results indicate that WMIs may be an effective treatment modality for patients with diagnosed mental health disorders. Future studies should evaluate the effects of clearly described, theory-driven, high-quality, and targeted WMIs in varied clinical populations via well designed and powered controlled trials. Finally, future research should determine patient as well as intervention variables that determine the take-up and therapeutic effects of WMIs.

Acknowledgments

The authors wish to thank investigators who provided additional information, including Dr Spijkerman, Dr Ly, Dr Johansson, Dr Dahlin, Dr Schenström, Dr Venmark, Dr Murray, and Dr Gerskovich.

Authors' Contributions

JSLJ performed the literature search and assessed all potentially relevant articles for inclusion. JSLJ and OSE independently assessed the retrieved articles, extracted relevant data, and rated each study's methodological quality. OSE performed statistical analysis. JSLJ wrote the manuscript's first draft, and OSE contributed the quantitative part on methods and results. IP critically revised the manuscript. JSLJ, MAJ and PM designed the study. MAJ and PM participated in the consensus process, and critically revised the manuscript. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Characteristics of the selected studies.

[PDF File (Adobe PDF File), 37KB - mental_v5i3e10278_app1.pdf]

Multimedia Appendix 2

Characteristics of the interventions.

[PDF File (Adobe PDF File), 33KB - mental_v5i3e10278_app2.pdf]

Multimedia Appendix 3

Depression. Type of therapy and component.

[PNG File, 38KB - mental_v5i3e10278_app3.png]

Multimedia Appendix 4

Depression. Type of control.

[PNG File, 33KB - mental_v5i3e10278_app4.png]

Multimedia Appendix 5

Depression. Pretreatment vs F-U data.

[PNG File, 31KB - mental_v5i3e10278_app5.png]

Multimedia Appendix 6

Anxiety. Type of therapy and component.

[PNG File, 36KB - mental_v5i3e10278_app6.png]

Multimedia Appendix 7

Anxiety. Type of control.

[PNG File, 34KB - mental_v5i3e10278_app7.png]

Multimedia Appendix 8

Anxiety. Pretreatment vs F-U data.

[PNG File, 30KB - mental_v5i3e10278_app8.png]

Multimedia Appendix 9

Quality of Life. Type of therapy and component.

[PNG File, 37KB - mental_v5i3e10278_app9.png]

Multimedia Appendix 10

Quality of life. Type of control.

[PNG File, 33KB - mental_v5i3e10278_app10.png]

Multimedia Appendix 11

Quality of life. Pretreatment vs F-U data.

[PNG File, 30KB - mental_v5i3e10278_app11.png]

Multimedia Appendix 12

Mindfulness. Pretreatment vs F-U data.

[PNG File, 27KB - mental_v5i3e10278_app12.png]

Multimedia Appendix 13

Publication bias of anxiety outcome.

[PNG File, 12KB - mental_v5i3e10278_app13.png]

Multimedia Appendix 14

Publication bias of depression outcome.

[PNG File, 15KB - mental_v5i3e10278_app14.png]

Multimedia Appendix 15

Publication bias of quality of life.

[PNG File, 12KB - mental_v5i3e10278_app15.png]

Multimedia Appendix 16

Publication bias of mindfulness outcome.

[PNG File, 11KB - mental_v5i3e10278_app16.png]

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493. [doi: 10.1093/ije/dyu038] [Medline: 24648481]
2. Kessler RC, Aguilar-Gaxiola S, Alonso J, Chatterji S, Lee S, Ormel J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. Epidemiol Psichiatr Soc 2009 Mar;18(1):23-33 [FREE Full text] [Medline: 19378696]
3. Kessler RC, Sampson NA, Berglund P, Gruber MJ, Al-Hamzawi A, Andrade L, et al. Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. Epidemiol Psychiatr Sci 2015 Jun;24(3):210-226. [doi: 10.1017/S2045796015000189] [Medline: 25720357]
4. Bebbington P, Brugha T, Coid J, Crawford M, Deverill C, D Souza J. Adult psychiatric morbidity in England Results of a household survey. London: NHS Information Centre for health and social care; 2009.
5. McDermott B, Baigent M, Chanen A, Fraser L, Graetz B, Hayman N. Clinical practice guidelines: depression in adolescents and young adults Internet. Melbourne: beyondblue; 2010.
6. National Institute for Health and Care Excellence (NICE). Generalised anxiety disorder and panic disorder in adults: management Internet (Clinical guideline [CG113]). 2011 Jan. URL: https://www.nice.org.uk/guidance/cg113 [accessed 2018-08-21] [WebCite Cache ID 71qCHTx2n]
7. NICE. Depression: The Nice Guideline on the Treatment and Management of Depression in Adults. Leicester: British Psychological Society & The Royal College of Psychiatrist; 2010.
8. Strauss C, Cavanagh K, Oliver A, Pettman D. Mindfulness-based interventions for people diagnosed with a current episode of an anxiety or depressive disorder: a meta-analysis of randomised controlled trials. PLoS One 2014;9(4):e96110 [FREE Full text] [doi: 10.1371/journal.pone.0096110] [Medline: 24763812]
9. Vøllestad J, Nielsen MB, Nielsen GH. Mindfulness- and acceptance-based interventions for anxiety disorders: a systematic review and meta-analysis. Br J Clin Psychol 2012 Sep;51(3):239-260. [doi: 10.1111/j.2044-8260.2011.02024.x] [Medline: 22803933]
10. NICE. Depression: the treatment and management of depression in adults. London, UK: Draft; 2009.

11. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of your Body and Mind to Face Stress, Pain and Illness. New York, NY: Delacorte; 1990.
12. Hayes S, Strosahl K. A Practical Guide to Acceptance and Commitment Therapy. New York: Springer; 2004.
13. Williams JMG. Mindfulness and psychological process. Emotion 2010 Feb;10(1):1-7. [doi: 10.1037/a0018360] [Medline: 20141295]
14. Godfrin KA, van HC. The effects of mindfulness-based cognitive therapy on recurrence of depressive episodes, mental health and quality of life: A randomized controlled study. Behav Res Ther 2010 Aug;48(8):738-746. [doi: 10.1016/j.brat.2010.04.006] [Medline: 20462570]
15. Teasdale JD, Segal ZV, Williams JM, Ridgeway VA, Soulsby JM, Lau MA. Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. J Consult Clin Psychol 2000 Aug;68(4):615-623. [Medline: 10965637]
16. Hayes S, Strosahl K, Wilson K. Acceptance and Commitment Therapy: An experiential approach to behavior change. New York: Guilford Press; 1999.
17. Khoury B, Lecomte T, Fortin G, Masse M, Therien P, Bouchard V, et al. Mindfulness-based therapy: a comprehensive meta-analysis. Clin Psychol Rev 2013 Aug;33(6):763-771. [doi: 10.1016/j.cpr.2013.05.005] [Medline: 23796855]
18. Chiesa A, Serretti A. Mindfulness based cognitive therapy for psychiatric disorders: a systematic review and meta-analysis. Psychiatry Res 2011 May 30;187(3):441-453. [doi: 10.1016/j.psychres.2010.08.011] [Medline: 20846726]
19. Piet J, Hougaard E. The effect of mindfulness-based cognitive therapy for prevention of relapse in recurrent major depressive disorder: A systematic review and meta-analysis. Clin Psychol Rev 2011:1032-1040.
20. Kuyken W, Warren FC, Taylor RS, Whalley B, Crane C, Bondolfi G, et al. Efficacy of Mindfulness-Based Cognitive Therapy in Prevention of Depressive Relapse: An Individual Patient Data Meta-analysis From Randomized Trials. JAMA Psychiatry 2016 Jun 01;73(6):565-574. [doi: 10.1001/jamapsychiatry.2016.0076] [Medline: 27119968]
21. Janssen L, Kan CC, Carpentier PJ, Sizoo B, Hepark S, Grutters J, et al. Mindfulness based cognitive therapy versus treatment as usual in adults with attention deficit hyperactivity disorder (ADHD). BMC Psychiatry 2015 Sep 15;15:216 [FREE Full text] [doi: 10.1186/s12888-015-0591-x] [Medline: 26373634]
22. Chiesa A, Serretti A. Are Mindfulness-Based Interventions Effective for Substance Use Disorders? A Systematic Review of the Evidence. Subst Use Misuse 2014;49(5):492-512. [doi: 10.3109/10826084.2013.770027]
23. Carmody J, Baer RA. How long does a mindfulness-based stress reduction program need to be? A review of class contact hours and effect sizes for psychological distress. J Clin Psychol 2009 Jun;65(6):627-638. [doi: 10.1002/jclp.20555] [Medline: 19309694]
24. McRoberts C, Burlingame G, Hoag M. Comparative efficacy of individual and group psychotherapy: A meta-analytic perspective. Gr Dyn Theory Res Pract 1998 Jun;2(2):101-117.
25. Tucker M, Oei T. Is Group More Cost Effective than Individual Cognitive Behaviour Therapy? The Evidence is not Solid Yet. Behav Cogn Psychother 2007;35(1):77-91.
26. Oh E, Jorm AF, Wright A. Perceived helpfulness of websites for mental health information: a national survey of young Australians. Soc Psychiatry Psychiatr Epidemiol 2009 Apr;44(4):293-299. [doi: 10.1007/s00127-008-0443-9] [Medline: 18830552]
27. Álvarez-Jiménez M, Gleeson JF, Bendall S, Lederman R, Wadley G, Killackey E, et al. Internet-based interventions for psychosis: a sneak-peek into the future. Psychiatr Clin North Am 2012 Sep;35(3):735-747. [doi: 10.1016/j.psc.2012.06.011] [Medline: 22929876]
28. Wang PS, Simon G, Kessler RC. The economic burden of depression and the cost-effectiveness of treatment. Int J Methods Psychiatr Res 2003;12(1):22-33. [Medline: 12830307]
29. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
30. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
31. NICE. Computerised cognitive behaviour therapy for depression and anxiety Internet (Technology appraisal guidance [TA97]). 2006 Feb 22. URL: https://www.nice.org.uk/guidance/ta97 [accessed 2018-08-21] [WebCite Cache ID 71qH1eoRz]
32. Spijkerman M, Pots W, Bohlmeijer E. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
33. Boggs JM, Beck A, Felder JN, Dimidjian S, Metcalf CA, Segal ZV. Web-based intervention in mindfulness meditation for reducing residual depressive symptoms and relapse prophylaxis: a qualitative study. J Med Internet Res 2014 Mar 24;16(3):e87-e41 [FREE Full text] [doi: 10.2196/jmir.3129] [Medline: 24662625]
34. Cavanagh K, Strauss C, Cicconi F, Griffiths N, Wyper A, Jones F. A randomised controlled trial of a brief online mindfulness-based intervention. Behav Res Ther 2013 Sep;51(9):573-578. [doi: 10.1016/j.brat.2013.06.003] [Medline: 23872699]

35. Hesser H, Gustafsson T, Lundén C, Henrikson O, Fattahi K, Johnsson E, et al. A randomized controlled trial of Internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. J Consult Clin Psychol 2012 Aug;80(4):649-661. [doi: 10.1037/a0027021] [Medline: 22250855]
36. Boettcher J, Aström V, Påhlsson D, Schenström O, Andersson G, Carlbring P. Internet-based mindfulness treatment for anxiety disorders: a randomized controlled trial. Behav Ther 2014 Mar;45(2):241-253 [FREE Full text] [doi: 10.1016/j.beth.2013.11.003] [Medline: 24491199]
37. Ly K, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 09;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
38. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [Medline: 19622552]
39. American Psychiatry Association (. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. Washington: American Psychiatric Press Inc; 2000.
40. World Health Organization (WHO). The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.
41. Shawyer F, Farhall J, Mackinnon A, Trauer T, Sims E, Ratcliff K, et al. A randomised controlled trial of acceptance-based cognitive behavioural therapy for command hallucinations in psychotic disorders. Behav Res Ther 2012 Feb;50(2):110-121. [doi: 10.1016/j.brat.2011.11.007] [Medline: 22186135]
42. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
43. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. URL: https://handbook-5-1.cochrane.org/ [accessed 2018-08-21] [WebCite Cache ID 71qIYYLXM]
44. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
45. Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, et al. Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. Behav Ther 2011 Dec;42(4):676-688. [doi: 10.1016/j.beth.2011.03.007] [Medline: 22035996]
46. Baum C, Kuyken W, Bohus M, Heidenreich T, Michalak J, Steil R. The psychometric properties of the Kentucky Inventory of Mindfulness Skills in clinical populations. Assessment 2010 Jun;17(2):220-229. [doi: 10.1177/1073191109356525] [Medline: 20040728]
47. Cardaciotto L, Herbert JD, Forman EM, Moitra E, Farrow V. The assessment of present-moment awareness and acceptance: the Philadelphia Mindfulness Scale. Assessment 2008 Jun;15(2):204-223. [doi: 10.1177/1073191107311467] [Medline: 18187399]
48. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Comprenhensive Meta-Analysis Software. Englewood: Biostat; 2006.
49. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Lawrence Earlbaum Associates; 1988.
50. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
51. Borenstein M, Hedges LV, Higgins J, Rothstein H. Introduction to Meta-Analysis. West Sussex, UK: John Wiley & Sons Inc; 2009.
52. Houghton V. A Quantitative Study of the Effectiveness of Mindfulness-Based Stress Reduction Treatment, Using an Internet-Delivered Self-Help Progran, for Women with Generalized Anxiety Disorder. Minneapolis: Capella University; 2008.
53. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K. Internet-based behavioral activation and acceptance-based treatment for depression: A randomized controlled trial Internet. J Affect Disord.? 2013:337.
54. Kivi M, Eriksson MCM, Hange D, Petersson E, Vernmark K, Johansson B, et al. Internet-based therapy for mild to moderate depression in Swedish primary care: short term results from the PRIM-NET randomized controlled trial. Cogn Behav Ther 2014;43(4):289-298 [FREE Full text] [doi: 10.1080/16506073.2014.921834] [Medline: 24911260]
55. Murray G, Leitan N, Berk M, Thomas N, Michalak E, Berk L, et al. Online mindfulness-based intervention for late-stage bipolar disorder: pilot evidence for feasibility and effectiveness. J Affect Disord 2015 Jun 01;178:46-51. [doi: 10.1016/j.jad.2015.02.024] [Medline: 25795535]
56. Dahlin M, Andersson G, Magnusson K, Johansson T, Sjögren J, Håkansson A, et al. Internet-delivered acceptance-based behaviour therapy for generalized anxiety disorder: A randomized controlled trial. Behav Res Ther 2016 Feb;77:86-95. [doi: 10.1016/j.brat.2015.12.007] [Medline: 26731173]
57. Gershkovich M, Herbert J, Forman E, Glassman L. Guided Internet-Based Self-Help Intervention for Social Anxiety Disorder With Videoconferenced Therapist Support. Cognitive and Behavioral Practice 2016 May;23(2):239-255. [doi: 10.1016/j.cbpra.2015.08.005]

58. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K. Internet-based behavioral activation and acceptance-based treatment for depression: A randomized controlled trial Internet. J Affect Disord.? 2013:337.
59. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided Acceptance and Commitment Therapy for social anxiety disorder and/or panic disorder provided via the Internet and a smartphone application: A randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
60. Johansson R, Björklund M, Hornborg C, Karlsson S, Hesser H, Ljótsson B, et al. Affect-focused psychodynamic psychotherapy for depression and anxiety through the Internet: a randomized controlled trial. PeerJ 2013;1:e102 [FREE Full text] [doi: 10.7717/peerj.102] [Medline: 23862104]
61. Strandskov SW, Ghaderi A, Andersson H, Parmskog N, Hjort E, Wärn AS, et al. Effects of Tailored and ACT-Influenced Internet-Based CBT for Eating Disorders and the Relation Between Knowledge Acquisition and Outcome: A Randomized Controlled Trial. Behav Ther 2017 Sep;48(5):624-637. [doi: 10.1016/j.beth.2017.02.002] [Medline: 28711113]
62. Geschwind N, Peeters F, Huibers M, van OJ, Wichers M. Efficacy of mindfulness-based cognitive therapy in relation to prior history of depression: randomised controlled trial. Br J Psychiatry 2012 Oct;201(4):320-325. [doi: 10.1192/bjp.bp.111.104851] [Medline: 22878133]
63. 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]
64. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
65. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
66. Alvarez-Jimenez M, Gleeson J, Bendall S, Rice S, D Alfonso S, Eleftheriadis D, et al. SU125. Momentum: A Novel Online Social Media, Mindfulness, and Strengths-Based Intervention to Promote Functional Recovery in Ultra High Risk (UHR) Patients. Schizophr Bull 2017 Mar;43(Suppl 1):206-207. [doi: 10.1093/schbul/sbx024.121]
67. Khoury B, Lecomte T, Gaudiano BA, Paquin K. Mindfulness interventions for psychosis: a meta-analysis. Schizophr Res 2013 Oct;150(1):176-184. [doi: 10.1016/j.schres.2013.07.055] [Medline: 23954146]
68. Crane RS, Brewer J, Feldman C, Kabat-Zinn J, Santorelli S, Williams JMG, et al. What defines mindfulness-based programs? The warp and the weft. Psychol Med 2017 Apr;47(6):990-999. [doi: 10.1017/S0033291716003317] [Medline: 28031068]
69. Chiesa A, Castagner V, Andrisano C, Serretti A, Mandelli L, Porcelli S, et al. Mindfulness-based cognitive therapy vs. psycho-education for patients with major depression who did not achieve remission following antidepressant treatment. Psychiatry Res 2015 Apr 30;226(2-3):474-483. [doi: 10.1016/j.psychres.2015.02.003] [Medline: 25744325]

Abbreviations

MBCT: Mindfulness-based Cognitive Therapy
MBI: Mindfulness-based interventions
MBSR: Mindfulness-based stress reduction
TAU: treatment as usual
WMIs: Web-based mindfulness interventions

Edited by J Torous; submitted 01.03.18; peer-reviewed by J Firth, N Chadi; comments to author 24.03.18; revised version received 14.05.18; accepted 16.05.18; published 25.09.18.

Please cite as:
Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M
Web-Based Mindfulness Interventions for Mental Health Treatment: Systematic Review and Meta-Analysis
JMIR Ment Health 2018;5(3):e10278
URL: http://mental.jmir.org/2018/3/e10278/
doi:10.2196/10278
PMID:30274963

Original Paper

A New Online Mental Health Training Program for Workplace Managers: Pre-Post Pilot Study Assessing Feasibility, Usability, and Possible Effectiveness

Aimée Gayed[1], BPsych (Hons), MCrim; Anthony D LaMontagne[2,3], MEd, MA, ScD; Allison Milner[2], BPsych (Hons), MEpi, PhD; Mark Deady[4], BPsych (Hons), PhD; Rafael A Calvo[5], PhD; Helen Christensen[4], BA (Hons), MPsych (Clin), PhD; Arnstein Mykletun[1,6,7,8,9], MSocio, MPsych, PhD; Nick Glozier[10], MA, MSc, MBBS, MRCPsych, FRANZCP, PhD; Samuel B Harvey[4], MBBS, MRCGP, MRCPsych, FRANZCP, PhD

[1]School of Psychiatry, Faculty of Medicine, University of New South Wales, Randwick, Australia
[2]School of Population and Global Health, The University of Melbourne, Melbourne, Australia
[3]Centre for Population Health Research, Deakin University, Geelong, Australia
[4]Black Dog Institute, Faculty of Medicine, University of New South Wales, Sydney, Australia
[5]School of Electrical and Information Engineering, University of Sydney, Sydney, Australia
[6]Department of Mental Health and Suicide, Norwegian Institute of Public Health, Oslo, Norway
[7]Department of Community Medicine,, University of Tromsø, Tromsø, Norway
[8]Centre for Work and Mental Health, Nordland Hospital Trust, Bodø, Norway
[9]Centre for Research and Education in Forensic Psychiatry and Psychology, Haukeland University Hospital, Bergen, Norway
[10]Brain and Mind Centre and Central Clinical School, School of Medicine, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

Corresponding Author:
Aimée Gayed, BPsych (Hons), MCrim
School of Psychiatry
Faculty of Medicine
University of New South Wales
Black Dog Institute Building
Hospital Road
Randwick, 2031
Australia
Phone: 61 +61293824517
Fax: 61 +61293828208
Email: a.gayed@unsw.edu.au

Abstract

Background: Mental health has become the leading cause of sickness absence in high-income countries. Managers can play an important role in establishing mentally healthy workplaces and coordinating their organization's response to a mentally ill worker.

Objective: This pilot study aims to evaluate the feasibility, usability, and likely effectiveness of a newly developed online training program for managers called *HeadCoach*. *HeadCoach* aims to build managers' confidence in supporting the mental health needs of staff and promote managerial behavior most likely to result in a more mentally healthy workplace.

Methods: In total, 66 managers from two organizations were invited to participate in this pre-post pilot study of *HeadCoach*, which was made available to managers to complete at their own pace over a 4-week period. Data were collected at baseline and post intervention via an online research platform. The difference in mean scores for each outcome between these two time points was calculated using paired samples t tests.

Results: Of all the invited managers, 59.1% (39/66) participated in the trial, with complete pre–post data available for 56.4% (22/39) of the participants. The majority of respondents reported positive engagement with the program. During the study period, managers' knowledge regarding their role in managing mental health issues (P=.01) and their confidence in communicating with employees regarding mental illness (P<.001) significantly increased. In addition, a significant increase was observed from the baseline in managers' self-reported actions to use strategies to prevent and decrease stress among their team members (P=.02).

Conclusions: Although caution is needed due to the absence of a control group, preliminary results of this study suggest that *HeadCoach* could be a feasible, acceptable, and efficient method of training managers in best workplace practices to help support the mental health needs of their staff.

(JMIR Ment Health 2018;5(3):e10517) doi:10.2196/10517

KEYWORDS

manager; supervisor training; workplace mental health; mental health education; online intervention; knowledge; attitudes; behaviour; eHealth

Introduction

In several high-income countries, mental health conditions have become the leading cause of long-term sickness absence and occupational incapacity [1-4]. The development or persistence of mental ill health for some workers might, in part, be related to their workplace [5], a link that has now been acknowledged as a major public health concern [4]. Anxiety and mood disorders are the most common mental illnesses reported in the working population [6-8]. Although treatable and often preventable, the rates of functional impairment due to psychiatric conditions within the working age group have increased over recent decades [9], which comes at a substantial cost to individuals, their workplaces, and, eventually, the economy [4,7,10-14]. Thus, there is a growing focus on elucidating how work can affect mental health and how it can be addressed through workplace-based mental health and well-being interventions [5,15,16].

Workplace mental health programs could provide an opportunity to alter modifiable risk and protective factors for mental health and a chance to aid the identification, treatment, and rehabilitation of workers with mental health problems. Some psychosocial working conditions have been recognized as primary sources of work-related stress, which, if not managed effectively, can adversely affect workers' well-being and productivity; these include conflicting and excessive work demands, a lack of job control, organizational failure to effectively communicate with staff, and poor collegial relationships and support [15,17]. Many of these workplace risk factors can be modified through decisions and adjustments, which managers are often in a position to make [18]. The degree to which the managers set a positive example of accepting attitudes and supportive behaviors toward the mental health of their staff can act as a protective factor for their workers. In addition, managers can react to mental ill health episodes in a way that could benefit the recovery process for workers [19]. Such strategies include, but are not limited to, facilitating regular conversations with an employee, maintaining a focus on an employee's well-being, and developing an appropriate return to work plan if a worker is on long-term sickness absence for a mental health issue, regardless of the underlying cause [20,21]. Overall, these various preventative and responsive strategies delivered across individual, team, and organizational levels can create a mentally healthy workplace that enhances the mental health of its employees [4]. Despite the availability of best practice guidelines detailing these primary, secondary, and tertiary approaches to managing workplace mental health, managers often report uncertainty with regard to how to best support employees experiencing or at risk of mental illness [7,21].

In order to address these concerns, many organizations are introducing training for managers in how to decrease work-based mental health risk factors for their employees, support their recovery, and facilitate successful return to work following a period of sickness absence for mental ill health. There is some evidence suggesting the value of specialized training delivered to managers to promote an understanding of the mental health needs of their workers and help increase managers' confidence in discussing mental health matters with their staff [2,3,22-24]. Further evidence supports that such manager training is effective in shifting stigmatizing attitudes regarding mental illness [24-27] and promoting the implementation of positive managerial behaviors to address mental health issues within their team [22,24,26] with an overall positive effect for manager training found across these outcomes [28]. Yet, evaluations of a selection of workplace-based mental health training programs have been unable to determine the beneficial effects on managers' attitude toward mental illness [18] or managerial behaviors of mental health issues either reported by managers themselves or objectively by their direct reports [18,29,30]. This disparity in outcomes may be due to the selection of components included in the training. It is becoming increasingly recognized that an integrated approach is considered the best practice in workplace mental health interventions [15]. An integrated approach incorporates strategies to prevent harm, promote positive mental health, and address mental health at the workplace irrespective of the cause of the illness [15]. However, the integration of these key components has yet to become standard in manager training, resulting in the dissemination of a series of uncoordinated educational programs.

This pilot study comprises the feasibility stage of the Medical Research Council (MRC) Framework for complex interventions [31] and elucidates the development and initial testing of the delivery of a comprehensive online training intervention for managers called *HeadCoach*. *HeadCoach* integrates the components recognized as key to mental health training and aims to build managers' confidence and ability to best support the mental health needs of the staff they supervise. The content for this online program has been derived from two separate face-to-face programs developed for managers with a focus on the mental health of their employees [22,32], which, when combined, encompass the recommended reactive and preventative components. We acknowledge that e-learning cannot offer some benefits that face-to-face contact with an educator may provide, and discussion with other participants in the course is often limited or not available. However, the

modification of the delivery format from face-to-face to a mobile-responsive website offers a more flexible, time-effective, and economical means of training a large number of staff members [33]. Participants have the opportunity to schedule training around the demands of their job and may also revisit the content within a standardized learning environment, providing a better opportunity for the consolidation of the course material.

This pilot study aims to test the feasibility and usability of the *HeadCoach* program with a small group of managers prior to evaluating the program as part of a larger randomized controlled trial (RCT). The objective data on program engagement, matched with participants' self-reported program rating scores and free text feedback, would provide information on the rates of adherence and user experience. Besides evaluating the uptake of and interaction with the program, this study aims to investigate the possible effectiveness of *HeadCoach* as a workplace mental health intervention for managers. We hypothesized that the *HeadCoach* program would help improve managers' self-reported confidence to respond to the needs of staff experiencing mental health issues and promote their implementation of managerial techniques that would create a more mentally healthy workplace. Although the results of this study would not be capable of determining the true effectiveness of the intervention due to the absence of a control group, findings from this pre–post design may provide an initial insight into participants' responses to and acceptability of the program and be a valuable first step in examining any impact that may be found from this type of training.

Methods

Development of Intervention

In accordance with the MRC Framework for complex interventions [31], we reviewed the relevant literature and a meta-analysis regarding manager training [28] to establish a theoretical basis for the development of the intervention. Following this, the *HeadCoach* online training intervention [34] was developed to improve workplace mental health by providing a flexible, easily accessible, and engaging training program for managers that informed them how to best aid the mental health needs of all their employees. Consistent with the concepts of the self-efficacy theory [35], which indicates that people are more likely to engage in particular behaviors when they feel more capable of attaining the desired behavior, this program aimed to build managers' confidence to effectively respond to the needs of employees experiencing mental health issues and implement evidence-based managerial styles that promote a more mentally healthy workplace environment.

Previously, collaborators on this trial have conducted cluster RCTs of two different face-to-face manager training programs. The "RESPECT" Manager Training Program [22] is a 4-hour training package delivered by a clinical psychologist or a consultant psychiatrist to small groups of managers; this package combined mental health knowledge and communication training to promote more appropriate reactive responses from managers when mental health issues arise in staff they supervise. A previous RCT of this manager training showed that it resulted in changes to managers' confidence and reactive behavior that lasted for six months at least [22]. The second existing face-to-face program, which was more preventative in its approach, aimed to improve psychosocial working conditions within an organization by changing the managers' behavior to best promote a mentally healthy workplace [36]; the principles within this program were based on the management competencies for preventing and reducing stress at work developed by the Health and Safety Executive (HSE) in the UK [37]. *HeadCoach* was developed by combining the content from these two face-to-face programs and transforming the material into a format compatible with online delivery. Educators, designers, and IT developers with experience in creating online mental health programs were consulted for the development and appearance of the learning platform. Their primary aim was to create a functional user experience that would appeal to the target audience; this included carefully considering aspects such as the style of language in which the content would be presented in and the visual design features selected for the branding.

Consultation Groups

As part of the modeling phase recommended when assessing complex interventions [31], we held meetings ahead of the pilot stage of the trial with representatives from various industries partnering on the evaluation of *HeadCoach*. Included in these meetings were members from the organizations' human resources, health and well-being, and media communication teams as well as managers representing potential users of *HeadCoach*. During these sessions, we consulted participants on the relevance of the vignettes to be included in the program as well as the appropriateness of the language and style. In addition, sections of the program were also shown to determine the feasibility and usability for different industry groups. Where appropriate, adjustments were made to the content and functionality based on the feedback received from the consultation groups.

HeadCoach Content

HeadCoach is a self-paced online intervention program that comprises three topics to be completed sequentially. Each topic contains a series of 10-minute modules featuring small sections of text, activities, short videos, practical activities, and topic summary exercises for individuals to complete. Textbox 1 presents the pilot program outline, and Figures 1-3 display the screen shots of various pages within *HeadCoach*. In addition, Supplementary File 1 details the vignettes included in the summary exercises for topics 2 and 3. The first topic, *common mental illnesses*, introduces mental health issues commonly found at the workplace. The following topic, topic 2, *how to help an employee you are concerned about*, outlines the signs managers can look out for to assist in recognizing people within their team who may be at risk of mental health issues, provides useful steps on how to initiate a conversation with employees who may be experiencing mental ill health, and discusses managerial techniques that can be implemented to support employees such as having regular catch ups with staff, creating a meaningful workplace where staff feel connected, and knowing what mental health resources are available for employees. In addition, it specifies ways to assist employees to stay at work

if they wish or return to work faster following a period of sickness absence owing to mental illness. Finally, topic 3, *minimizing mental health risks at the workplace*, aims to up-skill managers with techniques useful in altering a range of workplace mental health risk factors across the individual, team, and organizational levels [21] to create a mentally healthy workplace for their employees. Strategies include how to prevent and manage work-related stress for employees by being a respectful and responsible manager, manage and communicate existing and future work for staff, tailor the management of individuals based on their individual capabilities and needs, and proactively and objectively resolve difficult situations and conflict within the team. These techniques are based on observational data [21] and management standard frameworks provided by agencies such as the HSE in the UK [37].

Participants were expected to complete *HeadCoach* in approximately 2.5 hours, although it was designed in a way that users could work through the content at a pace that suits their learning style and job demands across a 4-week period with automated weekly reminders sent to prompt program adherence. In addition, the intervention was accessed online through a mobile-responsive website using a desktop, laptop, tablet, or mobile phone.

Study Design

This pilot study was primarily conducted to test the feasibility and usability of *HeadCoach* and investigate, through self-reported responses, early evidence of its possible effectiveness in altering managers' level of confidence regarding dealing with mental health matters at the workplace. In order to maximize the opportunity to test the functionality of the product and receive substantial feedback on the processes and quality of training, all participants were allocated to receive the intervention. Participants comprised managers employed by one of the two industry partners collaborating on the pilot study. *HeadCoach* was delivered to managers individually through their usual computer, tablet, or Web-based mobile phone. Through an online form, the managers provided informed consent to participate in the study.

Recruitment

Two organizations in Australia volunteered to participate in the pilot evaluation of *HeadCoach*. One organization was an equipment and machinery hire company servicing Australia across metropolitan, regional, and remote areas; managers in this organization supervised branches and employees across one or more offices or regions. The second organization provided a statewide fire and rescue service; managers in this organization who were eligible for this pilot study were employed at the duty commander or station officer level and were responsible for one or several different fire stations. For both organizations, the managerial level identified as relevant for this pilot study comprised supervisors who were the primary contact for staff members regarding periods of sickness absence or when workplace issues arose.

Textbox 1. The course outline for the pilot version of HeadCoach.

Topic 1: Common Mental Illnesses

- Module 1: Recognizing mental health issues
- Module 2: The workplace and its people
- Module 3: Topic summary exercises

Topic 2: How to help an employee you are concerned about

- Module 1: Identifying people at risk
- Module 2: Providing support
- Module 3: Having the talk
- Module 4: Facilitating help seeking
- Module 5: Modifying work to help recovery
- Module 6: Returning to work
- Module 7: Topic summary exercises

Topic 3: Minimizing mental health risks in the workplace

- Module 1: How to be a respectful and responsible manager
- Module 2: Managing and communicating existing and future work
- Module 3: Managing individuals within a team
- Module 4: Managing difficult emotions
- Module 5: Topic summary exercises

Figure 1. Screenshot of the *HeadCoach* dashboard.

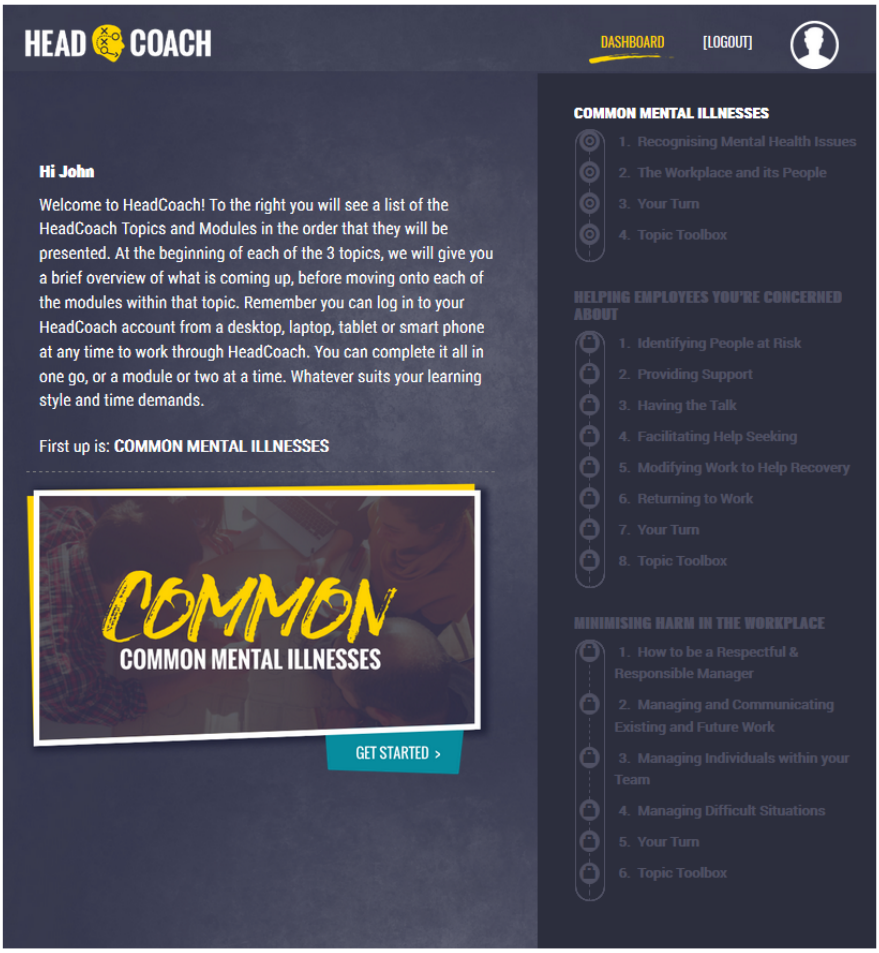


Figure 2. Screen shot of the landing page for a *HeadCoach* video resource.

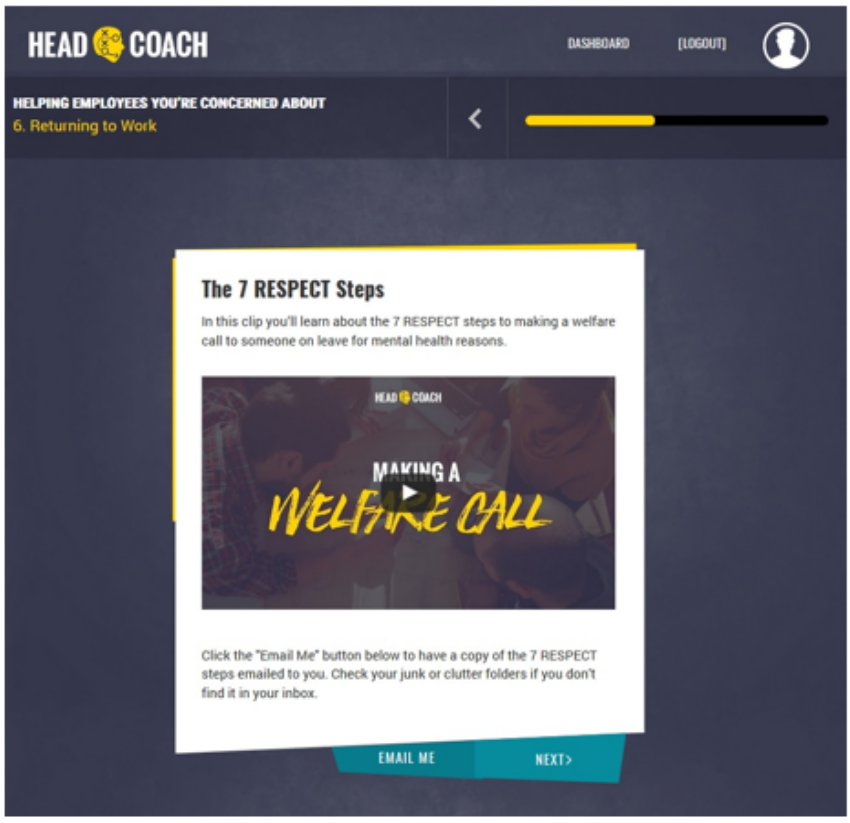
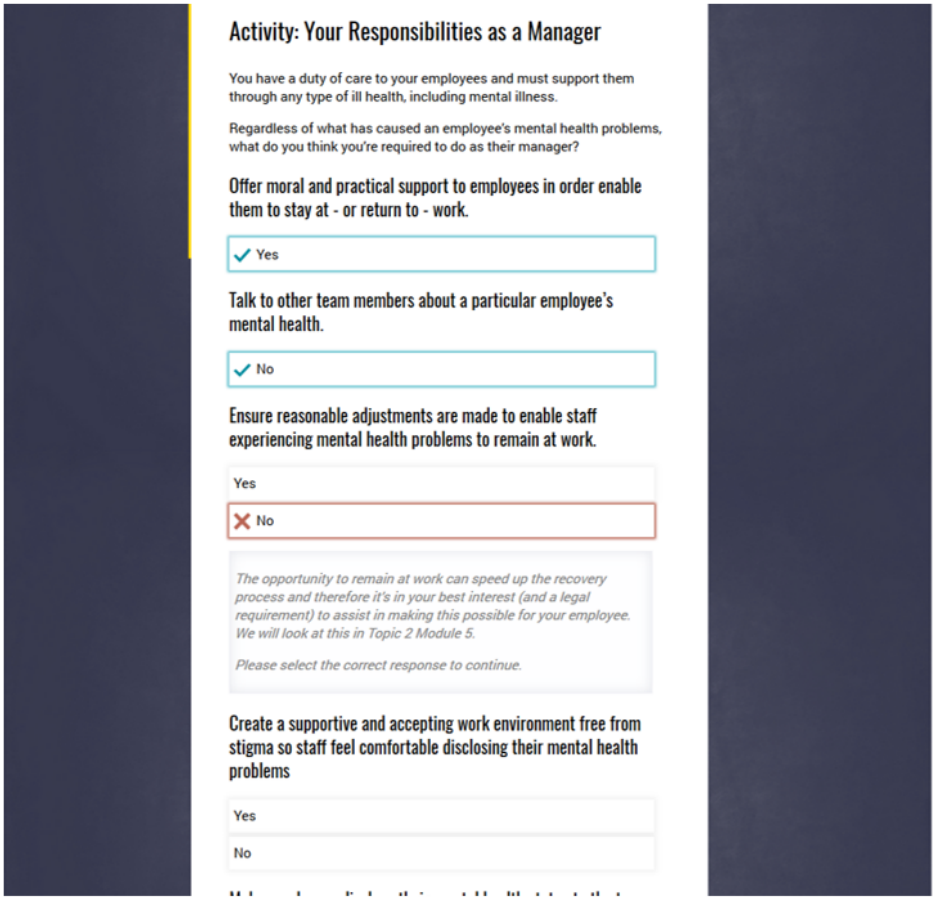


Figure 3. Screen shot of an interactive exercise with feedback in *HeadCoach*.



Participants met the inclusion criteria if they were currently supervising a team of ≥3 employees, aged ≥18 years, and currently residing in Australia with a good level of English comprehension. We contacted 66 managers via email who were identified by their employer as those fulfilling the inclusion criteria. The email described the purpose of the research, outlined what would be involved, and contained a link to register for the pilot study. In addition, the email emphasized that although participation in the trial was supported by their employer, it was entirely voluntary, and their level of involvement in the study would remain confidential from their employer.

Procedure

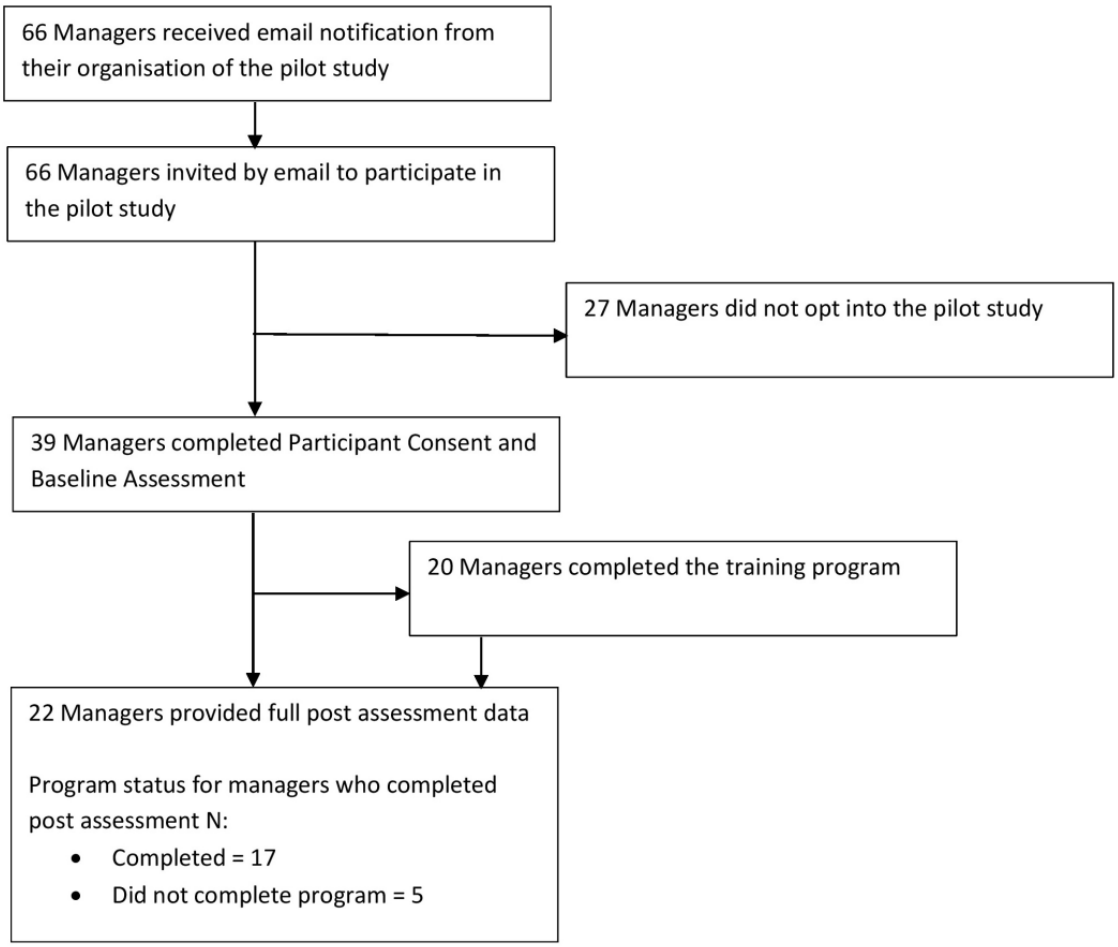
The trial procedure and stages of assessment for participants are outlined in Figure 4. All eligible to participate in the study received an information email from the researchers, inviting them to participate in the study. A hyperlink provided in the email directed them to the *HeadCoach* study website, which detailed further information about the program, the online consent form, and the registration page. Once registration was complete, the online baseline assessment was made available, following the completion of which participants were able to commence the online *HeadCoach* manager training program. The design of the program allowed managers to work through the program at their own pace over a 4-week period. Emails were automatically distributed weekly from sign up until participants had completed the program; these emails informed participants of the time remaining to complete the program and served as a reminder to revisit their account to continue working through the program.

Follow-up questionnaires were disseminated 4 weeks after completion of the baseline questionnaire. This 4-week period comprised the training period allocated to managers to complete *HeadCoach*. If all the components of the online program were completed earlier in the 4-week training period, the participant was invited to complete the follow-up questionnaire at that time point. This approach was selected to reduce nonresponse.

Data Collection

The baseline and postintervention data were collected electronically via the research platform that hosted the *HeadCoach* program, allowing a streamlined process between completing the *HeadCoach* program and completing questionnaires. In the case of nonresponse, two reminders were automatically sent across the subsequent 10 days via the research platform. On the completion of the 4-week postintervention questionnaire, the *HeadCoach* content was made available again for participants within their account for their reference.

Figure 4. *HeadCoach* pilot study procedure flow.



Measures of Usability and Feasibility

At the 4-week postintervention follow-up, participants were asked whether "This online course was engaging and interesting?" with response options ranging from *strongly disagree* to *strongly agree*. In addition, questions were asked evaluating the ease of navigating the program and finding information, whether the course fulfilled their expectations, and how likely would they recommend the program to their colleagues. Participants were also provided with an opportunity to provide detailed feedback through additional free text questions, including suggestions on what should be included in future versions. Finally, we obtained data regarding completion rates, including the duration of time taken to complete each module, from the research platform database.

Measures of Effectiveness

We measured managers' confidence in managing mental health issues and promoting a mentally healthy workplace using a modified version of the previously described supervisor scales [19]. Managers were asked to indicate their current level of confidence on a 5-point Likert scale with options ranging from *not at all confident* to *extremely confident*. Online Supplementary File 2 details the questions used for this measure. Knowledge about common mental health was assessed using the first 6 questions of the Mental Health Knowledge Schedule (MAKS) [38]. Nonstigmatizing attitudes toward mental illness was assessed using a modified version of previously published measures of personal stigma [39-41]. Comprehension of their role as a manager in dealing with mental health in the workplace was assessed using a questionnaire developed in accordance with the core competences outlined in a report detailing managers' role in supporting return to work after ill health [20]. The managers' application of managerial techniques that promote a mentally healthy workplace was assessed using an adapted version of the HSE Management Standards Indicator Tool [17]. We used a 5-point Likert scale with response options ranging from *strongly disagree* to *strongly agree* for questions such as "I provide regular opportunities for my team to speak one to one" and "I give employees the right level of job responsibility." Online Supplementary File 2 provides the remaining items. All questions were asked at the baseline and at the 4-week postintervention assessment. We converted the participants' scores to a percentage of the maximum possible score prior to data analysis.

We collected demographic information including age, gender, job role and length of service in the role, number of employees currently supervising, and previous mental health training.

Statistical Analysis

Participants' use of the training program was described in terms of the total time and number of modules completed. Descriptive statistics were used to demonstrate participants' responses to the questions on the usability and acceptability of *HeadCoach*. We assessed the differences in the mean percentage scores of each of the outcomes between baseline and post intervention collected at the 4-week follow-up using paired samples *t* tests. The parameters included managers' confidence in managing mental health issues within the workplace, level of mental health literacy, nonstigmatizing attitude toward mental illness, understanding of their role in managing mental health issues within their team, and application of managerial techniques that promote a mentally healthy workplace. All statistical analyses were conducted using SPSS version 23.

Results

Demographics

In this study, all managers were enrolled during November 2016 and December 2016 with the postintervention data collected during December 2016 and January 2017. Among the 66 managers who were invited to participate in the *HeadCoach* pilot study, 59.1% (39/66) registered and completed the baseline assessment. All managers were assigned to receive the intervention. Table 1 outlines the demographics of the study sample, including baseline characteristics of the entire sample and participants included in the final analysis.

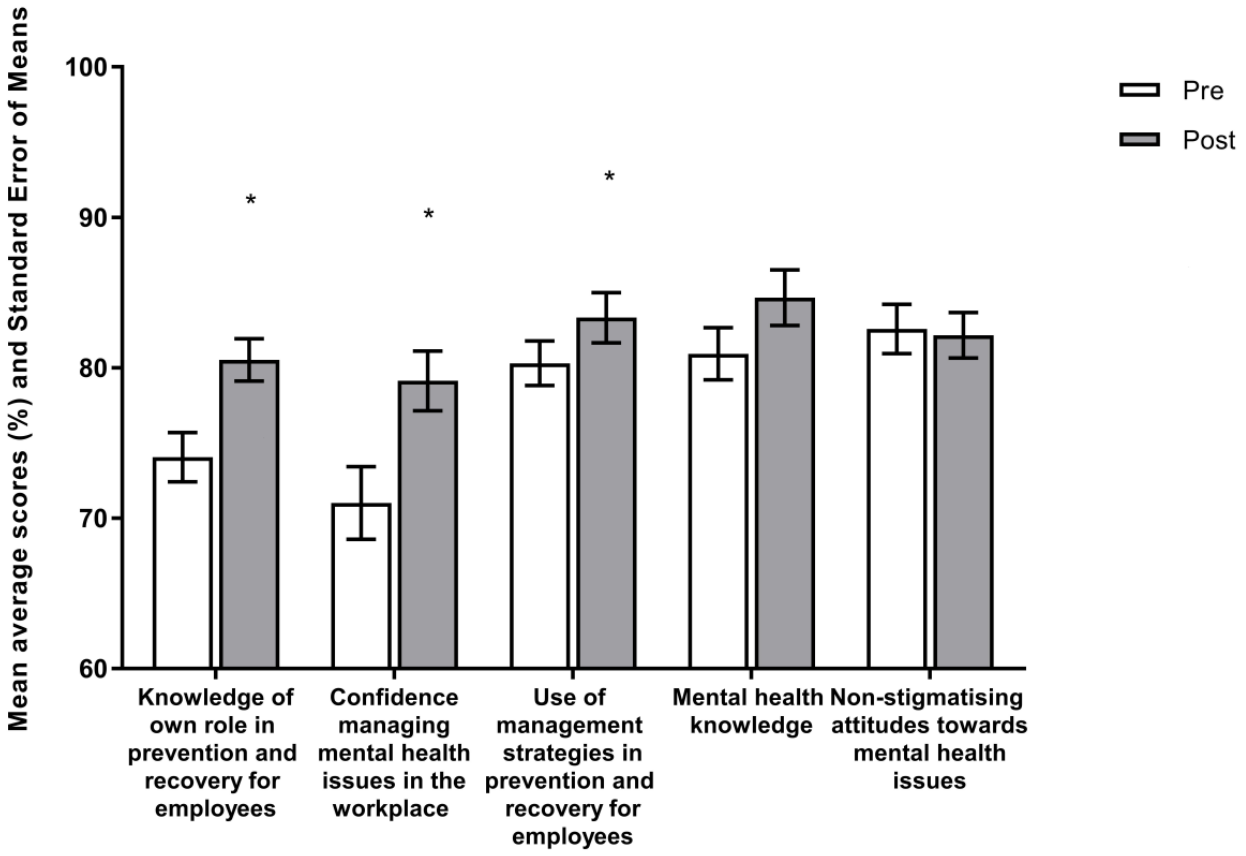
Usability and Feasibility of *HeadCoach*

Over half of the study sample (51.3%; 20/39) completed all 15 modules of the *HeadCoach* program. The time taken to complete the program ranged from 43 minutes to 2 hours 53 minutes. We obtained feedback on the program through the postquestionnaire. Of those who responded, 77% (17/22) found *HeadCoach* engaging and interesting, 86% (19/22) agreed it was easy to find the information needed, 73% (16/22) reported that the program fulfilled their expectations, 91% (20/22) considered it useful, and 86% (19/22) considered it worth recommending to a friend.

Table 1. Demographics as reported at the baseline.

Demographic	Total sample (n=39)	Sample with pre- and postintervention data available (n=22)
Age (years), mean (SD)	45.03 (8.74)	47.86 (8.44)
Gender, n (%)		
Male	34 (87.2)	20 (90.9)
Female	5 (12.8)	2 (9.1)
Industry, n (%)		
Building/Construction	12 (30.8)	5 (22.7)
Emergency Services	27 (69.2)	17 (77.3)
Years at current employer, n (%)		
<1	3 (7.7)	2 (9.1)
1-5	5 (12.8)	2 (9.1)
5-10	5 (12.8)	2 (9.1)
10-15	4 (10.3)	2 (9.1)
>15	22 (56.4)	14 (63.6)
Years in this level or above, n (%)		
<1	6 (15.4)	2 (9.1)
1-5	9 (23.1)	4 (18.4)
5-10	14 (35.9)	10 (45.5)
10-15	6 (15.4)	4 (18.2)
>15	4 (10.3)	2 (9.1)
Modules completed, n (%)		
0	7 (17.9%)	1 (4.5)
1-7	8 (20.5%)	1 (4.5)
8-14	4 (10.3%)	3 (13.6)
15 (Completed program)	20 (51.3%)	17 (77.3)
Mean (SD)	9.08 (6.62)	12.95 (4.37)

Figure 5. Baseline and 4-week post intervention mean average scores (%) and standard error of means for outcomes.



The qualitative feedback received post evaluation was primarily positive. All participants highly valued the practicality of the information and the format it was presented in. Examples of comments reflecting these views included "This topic was useful in giving logical and practical guidance on early intervention and strategies to minimize harm occurring in the workplace," "Good tips on initiating conversations and advice on where to go for resources," "It was great to get some practical examples and information that could be used in my workplace," and "Very well explained and easy to understand." In addition, comments around improved confidence to manage workplace mental health issues included "I found that it [*HeadCoach*] consolidated what I had covered previously in training and in doing so gave me more confidence to address an issue should one arise" and "[*HeadCoach*] gave me a bit more confidence in my approach." The key negative feedback regarding the content related to the level of detail included about common mental illnesses (eg, "Having a family history with mental illness, I hoped this course may be more informative, it tends to be a little superficial or shallow in details of symptoms to look out for"). This feedback prompted the implementation of more detailed information into the subsequent version of *HeadCoach* with the inclusion of direct access to additional online resources.

Effectiveness of *HeadCoach*

Figure 5 displays that following their use of *HeadCoach*, managers exhibited significantly higher levels of self-reported confidence in communicating with employees regarding mental illness (t_{22}=4.180; P<.001) and actions to employ managerial strategies to prevent and reduce stress among their team (t_{21}=2.468; P=.02). In addition, we observed a significant increase in managers' knowledge regarding their role in managing mental health issues (t_{23}=2.881; P=.01). At the follow-up, no significant change was noted in nonstigmatizing attitudes toward mental illness (t_{23}=–0.268; P=.80), whereas an increase in the levels of mental health knowledge fell just short of statistical significance (t_{24}=1.987; P=.06).

Discussion

Principal Findings

This pilot study assessed the feasibility, usability, and likely effectiveness of *HeadCoach*, a newly developed online training intervention for workplace managers regarding the mental health needs of employees reporting directly to them. To the best of our knowledge, *HeadCoach* is the first educational program for managers, which is delivered entirely online and provides an integrated program of preventative and reactive content consistent with recently recommended best practice frameworks for workplace mental health [4,15].

In this pilot study, we collected relevant data to assess the feasibility of recruitment and participation processes and the usability of the program through self-reported measures and also objectively, as provided through the online research platform. This online platform has the potential to provide valuable information regarding levels of adherence and engagement [42]. The findings suggest that for *HeadCoach*, the

online recruitment and registration processes, as well as methods for gathering data and delivering course material as operated through the current online research platform, were a practical and acceptable means for the dissemination and collection of various forms of information.

This study also suggests that *HeadCoach* correlated with significant increases in the managers' knowledge of their role in managing mental health issues among their staff, their confidence to do so, and their application of management strategies to promote a mentally healthy workplace. These changes in manager outcomes correspond to prior RCTs evaluating face-to-face training for managers promoting both reactive and preventative strategies to manage mental health issues within their team [22,26]. Although this study only evaluated the effect of *HeadCoach* with a pre–post design, the preliminary findings hold prominence because they suggest that e-learning options could replicate the positive outcomes previously reported with similar face-to-face manager training. This is an encouraging prospect because, if proven, it would allow organizations a practical means of training a large number of managerial staff about mental health issues with minimal time away from their jobs. Online training is a flexible and convenient format of learning because it can be structured to fit around other responsibilities and deadlines that managers are required to meet in their daily role. Although further evaluation through an RCT is warranted, these findings suggest the potential of *HeadCoach* to provide feasible, acceptable, and effective workplace mental health training for managers.

Limitations

Despite the novelty of our findings, there are a number of important limitations. As mentioned previously, the sample size was small with just over half of participants (51.3%; 20/39) completing the online intervention, and 56.4% (22/39) of the sample providing pre- and postintervention data. Although this dropout rate is typical in studies on online training [42,43] and survey responses [44], a higher response rate for both the adherence and follow-up would be ideal to minimize any potential bias caused by nonresponse. Besides the significance of personalized reminders [44], alternative engagement strategies may need to be adopted to help prompt adherence to the various stages of the trial, such as varying the messages contained within the email reminders and clearly conveying the purpose of participation [45]. In addition, strategies to maximize the response rates to the final questionnaire, such as the inclusion of a prize draw, may be valuable to increase the odds of participants responding [45].

The generalizability of results from this pilot study is limited by the opt-in approach to recruitment. The sample of managers who agreed to participate and who completed the multiple components of the trial might have held a pre-existing awareness of mental health issues with an interest to further develop their skills to best support their employees. This may explain the lack of substantial change found in managers' mental health knowledge because the sample may have already been well-informed of mental health issues and, therefore, a ceiling effect may have affected the outcomes of this study. Alternatively, the lack of marked effect may have been due to the lack of power, and with a larger sample size, a substantial change may have been observed for these outcomes. The generalizability of results may be limited by the sample of managers participating in this pilot study; the managers were from two specific industries, and nearly two-thirds (63.6%; 14/22) had been with their current employer for >15 years. An investigation within a variety of organizational contexts with a more representative sample is warranted before drawing conclusions about the feasibility and possible effects of this program for managers outside the building construction and emergency services industries.

The absence of a control group further limits the conclusions that could be drawn regarding the efficacy of the intervention because other factors explaining the observed change cannot be ruled out. Another limitation was the short follow-up period included in this trial. Although we observed a marked change postintervention, conclusions about the persistence of intervention effects over time cannot be drawn because of the lack of an extended follow-up. The short evaluation period included in this study may have also been insufficient to allow managers the opportunity to display a change in attitude within their work environment, yet with the inclusion of a longer follow-up, a change in nonstigmatizing attitudes could have been observed. A prolonged evaluation period would allow for employee-level data to be obtained for assessing the impact of manager training as perceived by the employees, providing valuable objective information about the flow on effects of manager training for the staff they supervise. The final potential limitation to note is the collection of self-reported data, especially regarding confidence levels. Participants may be more likely to report their levels of confidence or behavior favorably, although the use of identical questions at both time points and the anonymity of an online survey should have decreased this risk.

Conclusions

Although testing in a more comprehensive study with the inclusion of a control group is required to demonstrate a true effect, the preliminary findings from this pilot study suggest that *HeadCoach* may provide a practical and efficient method of training managers in best workplace mental health practices. Given the key role managers play in the promotion of mental health and well-being within their teams, there is great potential for this type of training to improve the support provided to employees for their mental health needs in the future.

Acknowledgments

This study was funded by *beyondblue* with donations from the Movember Foundation (Project Code: LK: 7139). Additional funding was provided by the icare foundation and the Mental Health Branch of NSW Health.

Authors' Contributions

AG and SBH conceptualized and designed the study. AG, ADL, AM, and SBH designed the intervention. AG acquired the data and with SBH, conducted the statistical analyses. AG and SBH drafted the manuscript. All coauthors contributed to the final manuscript.

Conflicts of Interest

HC and SBH are employed by the Black Dog Institute which provides manager training to workplaces.

Multimedia Appendix 1

Supplementary File 1 Vignettes for Summary Exercises.

[PDF File (Adobe PDF File), 147KB - mental_v5i3e10517_app1.pdf]

Multimedia Appendix 2

Supplementary File 2 Questionnaire.

[PDF File (Adobe PDF File), 62KB - mental_v5i3e10517_app2.pdf]

References

1. Harvey SB, Henderson M, Lelliott P, Hotopf M. Mental health and employment: much work still to be done. Br J Psychiatry 2009 Mar;194(3):201-203. [doi: 10.1192/bjp.bp.108.055111] [Medline: 19252144]
2. McLellan RK, Pransky G, Shaw WS. Disability management training for supervisors: a pilot intervention program. J Occup Rehabil 2001 Mar;11(1):33-41. [Medline: 11706775]
3. Tsutsumi A. Development of an evidence-based guideline for supervisor training in promoting mental health: literature review. J Occup Health 2011;53(1):1-9 [FREE Full text] [Medline: 21079372]
4. Petrie K, Joyce S, Tan L, Henderson M, Johnson A, Nguyen H, et al. A framework to create more mentally healthy workplaces: A viewpoint. Aust N Z J Psychiatry 2018 Jan;52(1):15-23. [doi: 10.1177/0004867417726174] [Medline: 28835112]
5. Harvey SB, Modini M, Joyce S, Milligan-Saville JS, Tan L, Mykletun A, et al. Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. Occup Environ Med 2017 Mar;74(4):301-310. [doi: 10.1136/oemed-2016-104015] [Medline: 28108676]
6. Harvey S, Henderson M. Occupational psychiatry. Psychiatry 2009 May;8(5):174-178 [FREE Full text] [doi: 10.1016/j.mppsy.2009.03.011]
7. Henderson M, Harvey SB, Overland S, Mykletun A, Hotopf M. Work and common psychiatric disorders. J R Soc Med 2011 May;104(5):198-207 [FREE Full text] [doi: 10.1258/jrsm.2011.100231] [Medline: 21558098]
8. van Zoonen K, Buntrock C, Ebert DD, Smit F, Reynolds CF, Beekman ATF, et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. Int J Epidemiol 2014 Apr;43(2):318-329 [FREE Full text] [doi: 10.1093/ije/dyt175] [Medline: 24760873]
9. Harvey SB, Deady M, Wang M, Mykletun A, Butterworth P, Christensen H, et al. Is the prevalence of mental illness increasing in Australia? Evidence from national health surveys and administrative data, 2001-2014. Med J Aust 2017 Jun 19;206(11):490-493. [Medline: 28918719]
10. Knudsen AK, Øverland S, Aakvaag HF, Harvey SB, Hotopf M, Mykletun A. Common mental disorders and disability pension award: seven year follow-up of the HUSK study. J Psychosom Res 2010 Jul;69(1):59-67. [doi: 10.1016/j.jpsychores.2010.03.007] [Medline: 20630264]
11. Knudsen AK, Harvey SB, Mykletun A, Øverland S. Common mental disorders and long-term sickness absence in a general working population. The Hordaland Health Study. Acta Psychiatr Scand 2013 Apr;127(4):287-297. [doi: 10.1111/j.1600-0447.2012.01902.x] [Medline: 22775341]
12. Harvey SB, Glozier N, Henderson M, Allaway S, Litchfield P, Holland-Elliott K, et al. Depression and work performance: an ecological study using web-based screening. Occup Med (Lond) 2011 May;61(3):209-211. [doi: 10.1093/occmed/kqr020] [Medline: 21525074]
13. Modini M, Joyce S, Mykletun A, Christensen H, Bryant RA, Mitchell PB, et al. The mental health benefits of employment: Results of a systematic meta-review. Australas Psychiatry 2016 Aug;24(4):331-336. [doi: 10.1177/1039856215618523] [Medline: 26773063]
14. Cocker F, Sanderson K, LaMontagne AD. Estimating the Economic Benefits of Eliminating Job Strain as a Risk Factor for Depression. J Occup Environ Med 2017 Jan;59(1):12-17. [doi: 10.1097/JOM.0000000000000908] [Medline: 28045792]
15. LaMontagne AD, Martin A, Page KM, Reavley NJ, Noblet AJ, Milner AJ, et al. Workplace mental health: developing an integrated intervention approach. BMC Psychiatry 2014 May 09;14:131 [FREE Full text] [doi: 10.1186/1471-244X-14-131] [Medline: 24884425]

16. Joyce S, Modini M, Christensen H, Mykletun A, Bryant R, Mitchell PB, et al. Workplace interventions for common mental disorders: a systematic meta-review. Psychol Med 2016 Mar;46(4):683-697. [doi: 10.1017/S0033291715002408] [Medline: 26620157]
17. Health Safety Executive. How to tackle work-related stress: A guide for employers on making the Management Standards work. www.hse.gov.uk. pdf: Health and Safety Executive, Executive HaS; 2009:430.
18. Nishiuchi K, Tsutsumi A, Takao S, Mineyama S, Kawakami N. Effects of an education program for stress reduction on supervisor knowledge, attitudes, and behavior in the workplace: a randomized controlled trial. J Occup Health 2007 May;49(3):190-198 [FREE Full text] [Medline: 17575399]
19. Nieuwenhuijsen K, Verbeek JHAM, de Boer AGEM, Blonk RWB, van Dijk FJH. Supervisory behaviour as a predictor of return to work in employees absent from work due to mental health problems. Occup Environ Med 2004 Oct;61(10):817-823 [FREE Full text] [doi: 10.1136/oem.2003.009688] [Medline: 15377767]
20. Chartered Institute of Personnel and Development. Report No 5244. London: CIPD; 2010 Jun. Manager support for return to work following long-term sickness absence: Chartered Institute of Personnel and Development June 2010 URL: https://www.cipd.co.uk/Images/manager-support-for-return-to-work-following-long-term-sickness-absence_2010_tcm18-10818.pdf [accessed 2018-06-13] [WebCite Cache ID 709zOeY3r]
21. Harvey S, Joyce S, Tan L, Johnson A, Nguyen H, Modini M, et al. Mental Health Commission, Australian Government. Australia: National Mental Health Commission and the Mentally Healthy Workplace Alliance; 2014 Nov. Developing a mentally healthy workplace: A review of the literature URL: http://www.mentalhealthcommission.gov.au/media/116414/Developing%20a%20mentally%20healthy%20workplace_Final%20November%202014.docx [accessed 2018-06-13] [WebCite Cache ID 709zdDmL8]
22. Milligan-Saville J, Tan L, Gayed A, Barnes C, Madan I, Dobson M, et al. Workplace mental health training for managers and its effect on sick leave in employees: a cluster randomised controlled trial. The Lancet Psychiatry 2017 Nov;4(11):850-858 [FREE Full text] [doi: 10.1016/S2215-0366(17)30372-3]
23. Takao S, Tsutsumi A, Nishiuchi K, Mineyama S, Kawakami N. Effects of the job stress education for supervisors on psychological distress and job performance among their immediate subordinates: a supervisor-based randomized controlled trial. Journal of occupational health Internet 2006 Nov;48(6):494-503 [FREE Full text]
24. Tsutsumi A, Takao S, Mineyama S, Nishiuchi K, Komatsu H, Kawakami N. Effects of a supervisory education for positive mental health in the workplace: a quasi-experimental study. J Occup Health 2005 May;47(3):226-235 [FREE Full text] [Medline: 15953844]
25. Moffitt J, Bostock J, Cave A. Promoting well-being and reducing stigma about mental health in the fire service. Journal of Public Mental Health 2014;13(2):103-113 [FREE Full text]
26. Dimoff JK, Kelloway EK, Burnstein MD. Mental health awareness training (MHAT): The development and evaluation of an intervention for workplace leaders. International Journal of Stress Management 2016;23(2):167-189. [doi: 10.1037/a0039479]
27. Shann C, Martin A, Chester A, Ruddock S. Effectiveness and Application of an Online Leadership Intervention to Promote Mental Health and Reduce Depression-Related Stigma in Organizations. J Occup Health Psychol 2018 Jan 04. [doi: 10.1037/ocp0000110] [Medline: 29300098]
28. Gayed A, Milligan-Saville JS, Nicholas J, Bryan BT, LaMontagne AD, Milner A, et al. Effectiveness of training workplace managers to understand and support the mental health needs of employees: a systematic review and meta-analysis. Occup Environ Med 2018 Mar 21. [doi: 10.1136/oemed-2017-104789] [Medline: 29563195]
29. Kawakami N, Takao S, Kobayashi Y, Tsutsumi A. Effects of web-based supervisor training on job stressors and psychological distress among workers: a workplace-based randomized controlled trial. Journal of Occupational Health 2006;48(1):28-34 [FREE Full text]
30. Stansfeld S, Berney L, Bhui K, Chandola T, Costelloe C, Hounsome N, et al. Pilot study of a randomised trial of a guided e-learning health promotion intervention for managers based on management standards for the improvement of employee well-being and reduction of sickness absence: the GEM (Guided E-learning for Managers) study. Public Health Research 2015;3(9). [doi: 10.3310/phr03090]
31. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655-a1655. [doi: 10.1136/bmj.a1655]
32. LaMontagne AD, Milner AJ, Allisey AF, Page KM, Reavley NJ, Martin A, et al. An integrated workplace mental health intervention in a policing context: Protocol for a cluster randomised control trial. BMC Psychiatry 2016 Feb 27;16:49 [FREE Full text] [doi: 10.1186/s12888-016-0741-9] [Medline: 26920745]
33. Kawakami N, Kobayashi Y, Takao S, Tsutsumi A. Effects of web-based supervisor training on supervisor support and psychological distress among workers: a randomized controlled trial. Prev Med 2005 Aug;41(2):471-478 [FREE Full text] [doi: 10.1016/j.ypmed.2005.01.001] [Medline: 15917043]
34. Gayed A, Bryan BT, Petrie K, Deady M, Milner A, LaMontagne AD, et al. A protocol for the HeadCoach trial: the development and evaluation of an online mental health training program for workplace managers. BMC Psychiatry 2018 Dec 29;18(1):25 [FREE Full text] [doi: 10.1186/s12888-018-1603-4] [Medline: 29378536]

35. Sweet SN, Fortier MS, Strachan SM, Blanchard CM, Boulay P. Testing a Longitudinal Integrated Self-Efficacy and Self-Determination Theory Model for Physical Activity Post-Cardiac Rehabilitation. Health Psychol Res 2014 Jan 13;2(1):1008 [FREE Full text] [doi: 10.4081/hpr.2014.1008] [Medline: 26973926]
36. LaMontagne A, Milner AJ, Allisey AF, Page KM, Reavley NJ, Martin A, et al. An integrated workplace mental health intervention in a policing context: Protocol for a cluster randomised control trial. BMC Psychiatry 2016 Feb 27;16:49 [FREE Full text] [doi: 10.1186/s12888-016-0741-9] [Medline: 26920745]
37. Yarker J, Lewis R, Donaldson-Feilder E. Management competencies for preventing and reducing stress at work: Identifying and developing the management behaviours necessary to implement the HSE Management Standards: Phase Two. Health and Safety Executive 2008 Jun:1-95 [FREE Full text]
38. Evans-Lacko S, Little K, Meltzer H, Rose D, Rhydderch D, Henderson C, et al. Development and psychometric properties of the Mental Health Knowledge Schedule. Can J Psychiatry 2010 Jul;55(7):440-448. [doi: 10.1177/070674371005500707] [Medline: 20704771]
39. Greenberg N, Gould M, Langston V, Brayne M. Journalists' and media professionals' attitudes to PTSD and help-seeking: A descriptive study. Journal of Mental Health 2009 Dec 16;18(6):543-548. [doi: 10.3109/09638230903191231]
40. Griffiths K, Crisp D, Jorm A, Christensen H. Does stigma predict a belief in dealing with depression alone? J Affect Disord 2011 Aug;132(3):413-417. [doi: 10.1016/j.jad.2011.03.012] [Medline: 21440305]
41. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: randomised controlled trial. Br J Psychiatry 2004 Oct;185:342-349 [FREE Full text] [doi: 10.1192/bjp.185.4.342] [Medline: 15458995]
42. 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]
43. Fear NT, Jones M, Murphy D, Hull L, Iversen AC, Coker B, et al. What are the consequences of deployment to Iraq and Afghanistan on the mental health of the UK armed forces? A cohort study. Lancet 2010 May 22;375(9728):1783-1797. [doi: 10.1016/S0140-6736(10)60672-1] [Medline: 20471076]
44. Cook C, Heath F, Thompson RL. A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. Educational and Psychological Measurement 2000 Dec 01;60(6):821-836. [doi: 10.1177/00131640021970934]
45. Sauermann H, Roach M. Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. Research Policy 2013 Feb;42(1):273-286 [FREE Full text] [doi: 10.1016/j.respol.2012.05.003]

Abbreviations

GEM: Guided e-learning for managers
MHAT: Mental health awareness training
MRC: Medical Research Council
RCT: Randomized controlled trial
HSE: Health and Safety Executive

Edited by G Eysenbach; submitted 28.03.18; peer-reviewed by J Wang, Q Yuan, S Pezaro; comments to author 09.05.18; revised version received 22.05.18; accepted 24.05.18; published 03.07.18.

Please cite as:
Gayed A, LaMontagne AD, Milner A, Deady M, Calvo RA, Christensen H, Mykletun A, Glozier N, Harvey SB
A New Online Mental Health Training Program for Workplace Managers: Pre-Post Pilot Study Assessing Feasibility, Usability, and Possible Effectiveness
JMIR Ment Health 2018;5(3):e10517
URL: http://mental.jmir.org/2018/3/e10517/
doi:10.2196/10517
PMID:29970359

Original Paper

Reaching Those At Risk for Psychiatric Disorders and Suicidal Ideation: Facebook Advertisements to Recruit Military Veterans

Alan R Teo[1,2,3], MD, MS; Samuel BL Liebow[2], BA; Benjamin Chan[3], MS; Steven K Dobscha[1,2], MD; Amanda L Graham[4,5], PhD

[1]HSR&D Center to Improve Veteran Involvement in Care (CIVIC), VA Portland Health Care System, Portland, OR, United States

[2]Department of Psychiatry, Oregon Health & Science University, Portland, OR, United States

[3]School of Public Health, Oregon Health & Science University and Portland State University, Portland, OR, United States

[4]Schroeder Institute for Tobacco Research & Policy Studies, Truth Initiative, Washington, DC, United States

[5]Department of Oncology, Georgetown University Medical Center, Washington, DC, United States

Corresponding Author:
Alan R Teo, MD, MS
HSR&D Center to Improve Veteran Involvement in Care (CIVIC)
VA Portland Health Care System
3710 SW US Veterans Hospital Road (R&D 66)
Portland, OR, 97239-2964
United States
Phone: 1 503 220 8262 ext 52461
Email: teoa@ohsu.edu

Related Article:

This is a corrected version. See correction statement: http://mental.jmir.org/2019/1/e13035/

Abstract

Background: Younger military veterans are at high risk for psychiatric disorders and suicide. Reaching and engaging veterans in mental health care and research is challenging. Social media platforms may be an effective channel to connect with veterans.

Objective: This study tested the effectiveness of Facebook advertisements in reaching and recruiting Iraq and Afghanistan-era military veterans in a research study focused on mental health.

Methods: Facebook ads requesting participation in an online health survey ran for six weeks in 2017. Ads varied imagery and headlines. Validated instruments were used to screen for psychiatric disorders and suicidality. Outcomes included impressions, click-through rate, survey completion, and cost per survey completed.

Results: Advertisements produced 827,918 impressions, 9,527 clicks, and 587 survey completions. Lack of enrollment in Veterans Affairs health care (193/587, 33%) and positive screens for current mental health problems were common, including posttraumatic stress disorder (266/585, 45%), problematic drinking (243/584, 42%), major depression (164/586, 28%), and suicidality (132/585, 23%). Approximately half of the survey participants (285/587, 49%) were recruited with just 2 of the 15 ads, which showed soldiers marching tied to an “incentive” or “sharing” headline. These 2 ads were also the most cost-effective, at US $4.88 and US $5.90 per participant, respectively. Among veterans with current suicidal ideation, the survey-taking image resulted in higher survey completion than the soldiers marching image (P=.007).

Conclusions: Facebook advertisements are effective in rapidly and inexpensively reaching military veterans, including those at risk for mental health problems and suicidality, and those not receiving Veterans Affairs health care. Advertisement image and headlines may help optimize the effectiveness of advertisements for specific subgroups.

(JMIR Ment Health 2018;5(3):e10078) doi:10.2196/10078

KEYWORDS

Facebook; social media; methodology; Veterans Affairs; veterans

Introduction

Military veterans who served during the Iraq and Afghanistan conflicts are at an elevated risk for a number of psychiatric problems. In a Veterans Affairs (VA) sample, one in four were found to have at least one mental health diagnosis, mostly commonly posttraumatic stress disorder (PTSD), depression, and alcohol and substance use disorders [1]. Heightened risk for suicide is also a major concern [2], with rates of suicide among veterans in the United States markedly higher than the general population, even after adjustment for age and gender differences [3]. Despite this, approximately 40% of Iraq and Afghanistan veterans have never accessed VA health services [4], and even when these veterans are in VA care, they may not be more inclined to utilize mental health care if experiencing active suicidal ideation [5].

Preventing veteran suicide has been, and continues to be, a top priority. Nonetheless, in-person health care appointments pose a significant barrier to Iraq and Afghanistan era veterans who tend to be younger and more likely to be employed than other veterans [6,7]. Other common barriers to formal help-seeking and treatment access, even in the presence of seemingly severe symptoms such as suicidal ideation or behaviors, include low perceived need [8-10], distance from health care facilities [11], and a desire to "handle the problem alone" [8]. In a sample of veterans who died by suicide in Oregon, an estimated 78% had not accessed VA health services [12].

Traditional strategies for recruiting participants into research studies can suffer from narrow reach, geographical limitations, costliness, and time-intensiveness. By comparison, recruitment via social media platforms, especially Facebook, may be faster, cheaper, and easier than traditional methods [13–15]. Among social media platforms, Facebook may be an especially important tool because it is the largest and most used, with a diverse base of users with detailed demographic profiles [13]. Facebook users spend upwards of 50 minutes a day on the platform, and among online adults between the ages of 18 to 29, approximately 9 of 10 use Facebook [14,15].

In recent years, many studies have examined Facebook advertisements (ads) as a recruitment method for research studies on mental or behavioral health. Facebook ads have been used effectively in populations including college students and young adults [16,17] and military veterans particularly those with risky drinking [20–23]. Prior research has also suggested ad campaigns can achieve both "broad reach and targeted recruitment," and found ad costs to be manageable [18]. Throughout this paper, we use the term "reach" to mean even a minimal amount of engagement with an ad, not necessarily engagement in health care, a definition that is consistent with the vein in which it is used in social media contexts. Nonetheless, much work remains in the development of best practices and evidence-based recruitment strategies on social media. Studies conducting experiments comparing particular Facebook advertising approaches, such as differing images or text are lacking [19,21,25–27]. One recent study by Pedersen and colleagues, which focused on recruiting spouses of heavy drinking service members and veterans, did systematically and sequentially test different ad features. It concluded that ads with text accentuating the US $120 financial incentive for study participation had a higher conversion to study participation at a lower per-participant cost [19]. Examination of a spectrum of potential outcomes, ranging from general exposure (eg, "impressions") to initial interaction (eg, link clicks) to implementation (eg, enrollment in a research study) [20], would also be helpful in social media studies.

We conducted a study with three aims: (1) determine the feasibility of reaching Iraq and Afghanistan era military veterans through Facebook ads, (2) quantify the extent to which reached veterans are at risk for psychiatric problems, and (3) characterize how veterans utilize social media and interact with their social networks on Facebook. In this article, we focus on the first two aims. More specifically, we determined the recruitment of military veterans to a mental health focused research study, examined what ad features are most relevant to engaging veterans, and characterized what kinds of veterans are likely to be reached by the ads. As an exploratory study, we limited our a priori hypothesis to state that it would be feasible to recruit recent military veterans with probable mental health problems.

Methods

Participants

The target population for the survey was United States (US) military veterans of the Operation Enduring Freedom-Operation Iraqi Freedom (OEF-OIF) service era (September 2001-present), also referred to as Iraq and Afghanistan era veterans. To be eligible, individuals needed to be age 18 or older, and on active duty in the US Armed Forces after September 2001 but not presently. We excluded individuals who completed surveys in less than five minutes, had more than one survey response, or incorrectly answered a military-related "insider knowledge" question to reduce the chance of online survey misrepresentation [18,21]. Survey completion was defined as those respondents who reached the end of the online survey and were not excluded based on the above quality control measures.

Advertisement Campaign

Facebook offers myriad options related to placement and targeting of ads; the same parameters were used for all ads. Ads were run simultaneously, to identical audiences, with the same ad budget, and for the same duration of time. Ads ran for a total of 45 days between January 13, 2017, and March 18, 2017, except for one ad that was briefly deactivated by Facebook for technical reasons. Ads were exclusively placed in the News Feed on computers and not mobile phones as the survey was developed for computer administration. Ads were optimized per Facebook's algorithm for clicks, meaning that ads were automatically shown to users whom Facebook anticipated would click at the highest rates, in a targeting process adjusted by actual clicks during the campaign.

Study ads were broadly targeted at Facebook users in the US of any age or gender who had at least one of a variety of veteran-related characteristics (eg, interest in "United States Armed Forces" or "Supporting Our Veterans" as determined by their Facebook profile). Text above the ad image indicated

that veterans who served between 2001 and 2017 were needed for an "online health survey."

We designed a total of 15 ad variations in a 3x5 factorial design, with 3 images (ie, a person taking a survey on a tablet device; a veteran with his family; and soldiers marching) varied against five headlines. Ad images are illustrated in Figure 1. Headlines were informed by empirical research in psychology and survey methodology as well as established principles in behavioral economics known to help nudge behavior [22].

Figure 1. Sample Facebook ads illustrating the three different ad images (Survey-Taking, Family, and Soldiers Marching).

Specific approaches that can motivate research participation include: targeting feelings of altruism or a "warm-glow" effect [23], harnessing "psychological capital", which is closely correlated with a sense of empowerment [24,25], or using a statement of what others in similar situations do, also known as descriptive social norms [26]. Providing financial incentives [27] or encouraging the sharing of content with social network members could also increase engagement [28]. Based on these ideas, we crafted and used the following 5 ad headlines:

1. *Altruism*: "Will you help us improve care for veterans?"
2. *Empowerment*: "You can tell us how to design new health programs for veterans."
3. *Social norms*: "Hundreds of veterans are participating in this survey. Will you join them?"
4. *Incentive*: "You can win a new 7.9" 16 GB iPad Mini 4 with Retina Display!"
5. *Sharing*: "Will you share this with one veteran you know?"

Ads were hosted by Oregon Health & Science University (OHSU) and linked to an online survey. To calculate survey participation and other outcomes by an ad, we constructed separate uniform resource locators (URLs) to the online survey for each ad. Prospective participants initially completed an online consent and eligibility screener. As an incentive for survey completion, we informed potential participants of an optional sweepstakes or lottery, in which two randomly selected survey participants who provided contact information would receive an iPad. Eligible, consented participants then completed the full online survey. Before and after survey, we provided all participants with a series of online, phone, and text messaging mental health treatment resources, including options for crisis situations, non-urgent treatment referral, and—as we would not be aware of the particular location of respondents—ways to locate local support and treatment resources. The institutional review board of OHSU approved all study procedures.

Measures

Our primary outcome was survey completion, which represents the highest level of engagement with a Facebook ad [20]. As additional outcomes, we included measures of ad engagement that are automatically tabulated by Facebook for advertisers:

- *Impressions*: the total number of times that the ad is presented to any Facebook user.
- *Clicks*: the number of times that a user clicks on the ad.
- *Click-through rate (CTR)*: the number of clicks divided by impressions.
- *Reactions*: the total number of "Likes" or other Facebook reactions ("Love," "Haha," "Wow," "Sad," and "Angry") generated by an ad.

The online survey included a series of self-report questions to obtain background information about the sample including sociodemographic characteristics, military history, social media use, and interest in social media-based interventions. Using survey items from the National Survey of Veterans [29], we assessed the period of service (ie, "Have you ever served on active duty in the US Armed Forces? *Active duty includes serving in the US Armed Forces as well as activation from the Reserves or National Guard." and* " When did you serve on active duty in the US. Armed Forces? *Mark all service eras that apply."*), branch of service (ie, "In which branch or branches did you serve on active duty?") and deployment history (ie, "Did you deploy in support of Operation Enduring Freedom (OEF) or Operation Iraqi Freedom (OIF) or Operation New Dawn (OND)") Frequency of Facebook use was assessed by adapting previously validated survey items used by Pew Research [30]. We used two additional items from the National Survey of Veterans to determine VA health service use (ie, "Have you ever been enrolled in VA health care?" and "In the past 12 months, did you use any VA health care services?") [29]. Participants who responded "Don't know" to health service use questions were classified as not non-users.

To screen for mental health problems, we employed reliable and valid self-report tools. For PTSD, we used the Primary Care PTSD Screen for DSM-5 (PC-PTSD), a five-item scale assessing past-month symptoms of a lifetime traumatic event. A score of three or higher on the PC-PTSD indicates a positive screen [31]. For alcohol misuse, we used the Alcohol Use Disorders Identification Test Alcohol Consumption Questions (AUDIT-C), a three-item scale on frequency and intensity of drinking. An AUDIT-C score of four or higher for men, or three or higher for women, indicates a positive screen for problematic drinking [32]. For major depression, we used the Patient Health Questionnaire-2 (PHQ-2), a two-item scale on anhedonia and depressed mood in the previous two weeks. A score of two or higher on the PHQ-2 indicates a positive screen [33]. For suicidality, we used the Depressive Symptom Inventory Suicidality Subscale (DSI-SS), a four-item scale on suicidal ideation within the past two weeks [34]. A score of two or higher on the DSI-SS indicates a positive screen in a population-based sample [35].

Statistical Analysis

Demographic variables were compared by ad text and image for participants in the analytic sample using Pearson's chi-square test, or ANOVA for age. All outcomes were modeled as negative binomial counts. The study design parameters of image, headline, and the interaction were included as independent factors. The model for clicks and CTR included an offset for the number of impressions; the model for reactions included an offset for the number of clicks.

Results

Feasibility of Recruiting Military Veterans Through Facebook Advertisements

Over the 45 days of the advertising campaign (Figure 2), the Facebook ads produced 827,918 impressions, 9,527 clicks (CTR=9,527/827,918, 1.20%), and 1,787 reactions. There were 1,329 complete responses to the eligibility screener, of which 711 met eligibility criteria, and 605 completions of the online survey (ie, 605/711, 85% response rate). A total of 18 responses were excluded from the analysis based on quality control measures. Ten took less than 5 minutes, 2 claimed nonexistent pay grades, and 6 were duplicate responses. This left a final sample of 587 (ie, roughly 13 new participants each day). Total ad expenditure was US $11,427, yielding an average cost per analyzed survey of US $19.47.

Characteristics of Recruited Veterans

Characteristics of survey participants are described in Table 1. Their mean age was 40 years. A total of 81% (474/587) were male, and 81% (477/587) were white and non-Hispanic. In addition to serving during the Iraq and Afghanistan era, many also served during prior eras, particularly the Gulf War era (213/587, 36%). The majority (426/587, 73%) had been deployed in support of OEF-OIF. In this sample, the majority (326/587, 56%) were in the Army, as compared with approximately 36% among active-duty military personnel [36].

With regards to VA enrollment, 33% of participants (193/587) had not enrolled in VA health services and 55% (322/587) had not used VA care in the prior year. Positive screens for current mental health problems were common: PTSD (266/585, 45%), problematic drinking (243/584, 42%), and major depression (164/586, 28%). Twenty-three percent (132/585) indicated current suicidality. Of the participants not enrolled in VA health services, 21% (40/193) reported current suicidality.

Associations Between Advertisement Characteristics and Demographic Characteristics

Gender of respondents varied by ad text, (χ^2_4 (N=583)=10.7, *P*=.03), with *sharing* and *empowerment* messages having a higher proportion of women. Age varied by text (F(2, 585)=11.84, *P*<.01) and image (F(2, 585)=10.09, *P*<.01), with the *soldiers marching* image and *incentive* text attracting the youngest respondents and the *survey-taking image* and *altruism* and *social norms* headlines attracting the oldest. Service era varied correspondingly, with the *soldiers marching* image (χ^2_2 (N=585)=9.90, *P*<.01) and *incentive* text (χ^2_4, N=585)=22.03, *P*<.01) attracting a higher proportion of respondents who had only served during the OEF-OIF era. Race, ethnicity, education, military branch, and deployment to Iraq or Afghanistan did not vary significantly by ad text or image.

Variations in Advertisement Engagement and Cost

There was a main effect for ad image across impressions, CTR, and reactions, but not survey participation. In terms of both impressions and click-through rates, the *soldiers marching* image performed better than the *survey-taking* and *family* images (*P*<.001 for all comparisons). In addition, the *soldiers marching* image generated significantly more reactions than the *survey-taking* (*P*=.001) and *family* (*P*<.001) images. However, there were no significant differences by ad image in terms of survey participation.

Figure 2. Cumulative ad campaign outcomes over time.

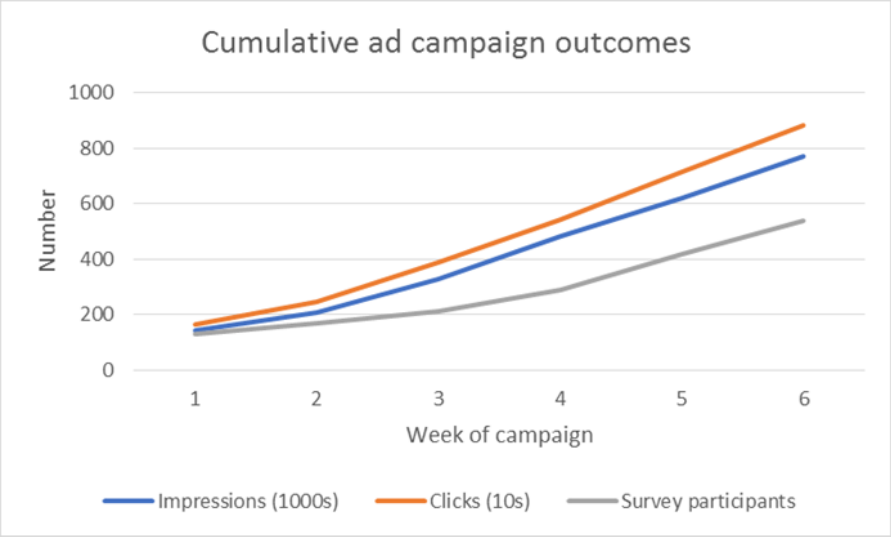


Table 1. Descriptive characteristics of all survey participants (N=587).

Characteristic	Value
Demographics and Military History	
Age in years, mean (SD)	40.0 (12.0)
Male, n (%)	474 (80.8)
Racial or ethnic minority, n (%)	110 (18.9)
Branch of military service[a], n (%)	
Army	326 (55.5)
Navy	110 (18.7)
Air Force	109 (18.6)
Marine Corps	75 (12.8)
Coast Guard	11 (1.9)
Other	2 (0.3)
Service era[b], n (%)	
September 2001-present (includes Iraq and Afghanistan conflicts)	587 (100)
August 1990-August 2001 (includes Gulf War)	213 (36.3)
May 1975-July 1990	112 (19.1)
August 1964-April 1975 (includes Vietnam era)	29 (4.9)
Deployed to Iraq or Afghanistan	426 (72.7)
Education, n (%)	
High school diploma or less	34 (5.8)
Some college, or vocational degree	250 (42.6)
College degree or greater	303 (51.6)
Marital status, n (%)	
Single, never married	112 (19.1)
Divorced, separated, or widowed	111 (18.9)
Married or living as married	363 (62.0)
Facebook use frequency, n (%)	
Every few weeks or less often	14 (2.4)
Weekly or a few times a week	47 (8.0)
Daily or more often	524 (89.6)
Clinical characteristics, n (%)	
Positive depression screener[c]	164 (28.0)
Positive posttraumatic stress disorder screener[d]	266 (45.5)
Positive alcohol misuse screener[e]	243 (41.6)
Positive suicidal ideation screener[f]	132 (22.6)
Veterans Affairs Health Service Use, n (%)	
Not enrolled	193 (32.9)
Not used in last year	322 (54.9)

[a]Percentages do not add up to 100% because some respondents (45/587, 7.7%) indicated multiple branches.

[b]Percentages do not add up to 100% because some respondents indicated multiple service eras; 60.5% (355) served only during the current era, September 2001 to present.

[c]Patient Health Questionnaire-2 (PHQ-2) score ≥ 3. Due to missing item response, the total number of respondents for this scale was 586.

[d]Primary Care posttraumatic stress disorder (PTSD) Screen for DSM-5 (PC-PTSD-5) score ≥ 3. Due to missing item response, the total number of

respondents for this scale was 502.

[e]Alcohol Use Disorders Identification Test Alcohol Consumption Questions (AUDIT-C) score ≥ 4 (men) or ≥ 3 (women). Due to missing item response, the total number of respondents for this scale was 477.

[f]Depressive Symptom Inventory Suicidality Subscale (DSI-SS) score ≥ 2. Due to missing item response, the total number of respondents for this scale was 585.

Table 2. Matrix of 15 Facebook advertisement variants with outcomes for each advertisement.

Advertisement Headline Type	Advertisement Image Type					
	Survey-Taking		Family		Soldiers Marching	
	Number (CTR[a])	Cost (US $)	Number (CTR)	Cost (US $)	Number (CTR)	Cost (US $)
Altruism	18 (0.82)	42.31	9 (1.11)	84.61	26 (1.43)	29.29
Empowerment	4 (0.87)	190.38	16 (0.99)	47.59	25 (1.12)	30.46
Incentive	20 (0.73)	38.08	20 (0.85)	38.08	156 (1.27)	4.88
Sharing	44 (0.97)	17.31	26 (1.06)	29.29	129 (1.64)	5.90
Social Norms	22 (1.10)	34.61	13 (0.99)	58.58	57 (1.81)	13.36

[a]CTR: click-through rate.

There was also a main effect for ad headline on ad engagement outcomes. Specifically, the *sharing* headline was associated with more impressions than the *incentive* (P=.045) and *empowerment* (P=.004) headlines; more reactions than the *altruism* (P=.004) and *empowerment* (P=.014) headlines; and higher survey participation than the *social norms* headline (P<.001). In addition, the *social norms* headline was associated with higher click-through rates than the *incentive* (P<.001), *altruism* (P<.001), and *empowerment* P=.001) headlines.

Two of the 15 ad versions generated nearly half (285/587, 49%) of the participants (Table 2). These were the ads containing the image of *soldiers marching* with either the *incentive* or *sharing* headline. These two ad versions had significantly higher impressions (P<.001) and CTR (P<.001) than the other 13. Consequently, these two ads were most cost effective, at US $4.88 and US $5.90 per participant, respectively. Results were similar when examining individuals who completed the online eligibility screener, regardless of whether they were eligible or completed the full survey.

Veterans With Suicidal Ideation or Non-Enrolled in Veterans Affairs Health Care

Among veterans who completed the survey, the probability of suicidal ideation ranged from an estimated 15%-50% across the 15 ad variants, and the probability of not being enrolled in VA health care ranged from an estimated 18%-50% (Table 3). Recruitment of veterans with suicidal ideation was significantly higher for ads with the *survey-taking* image, as compared to the *soldiers marching* (P=.007) image. There were no statistically significant differences in recruitment of non-enrolled veterans by ad image or headline.

Table 3. Predicted probabilities of suicide ideation and non-enrollment in Veterans Affairs Health Care among survey participants.

Advertisement image and headline	Probability of suicidal ideation[a] (95% CI)	Probability of not being enrolled in veterans affairs health care (95% CI)
Survey-taking		
Altruism	0.28 (0.12-0.52)	0.28 (0.12-0.52)
Empowerment	0.25 (0.03-0.76)	0.25 (0.03-0.76)
Incentive	0.45 (0.25-0.66)	0.50 (0.29-0.71)
Sharing	0.32 (0.20-0.47)	0.30 (0.18-0.44)
Social norms	0.19 (0.07-0.41)	0.18 (0.07-0.40)
Family		
Altruism	0.22 (0.06-0.58)	0.33 (0.11-0.67)
Empowerment	0.50 (0.27-0.73)	0.25 (0.10-0.51)
Incentive	0.20 (0.08-0.43)	0.30 (0.14-0.53)
Sharing	0.15 (0.06-0.35)	0.38 (0.22-0.58)
Social norms	0.23 (0.08-0.52)	0.38 (0.17-0.66)
Soldiers Marching		
Altruism	0.19 (0.08-0.39)	0.23 (0.11-0.43)
Empowerment	0.16 (0.06-0.36)	0.28 (0.14-0.48)
Incentive	0.17 (0.12-0.24)	0.37 (0.30-0.45)
Sharing	0.22 (0.15-0.30)	0.33 (0.25-0.41)
Social norms	0.25 (0.15-0.37)	0.33 (0.22-0.46)

[a]Depressive Symptom Inventory Suicidality Subscale (DSI-SS) score ≥ 2.

Discussion

Key Findings

Our study demonstrated that Facebook ads are a potentially powerful tool to recruit research subjects. With the support of a single, half-time research assistant, we engaged veterans in enrollment in our online survey at a rapid clip (ie, nearly 100 participants per week). Though our click-through rate was similar to prior studies, our response rate was very high, which may have reflected ease of participation in this online survey. Average cost per participant was less than US $20, and our best-performing ads were dramatically cheaper, approximately US $5-6 per survey participant, a figure that compares very favorably with most prior studies [37,38]. Facebook ads were further disseminated through social sharing, as is illustrated by the "likes", comments, and sharing of ads that we observed. This is a significant "externality" from a cost efficiency standpoint.

The feasibility of reaching and engaging younger veterans in research through this approach has important public health and clinical implications. We reached not only a relatively broad target population (ie, recent military veterans) but also were effective in engaging subpopulations that can be hard-to-reach and are of heightened interest. Being able to rapidly reach veterans who are experiencing current suicidal ideation and unengaged in VA health care—as we did—is a major challenge for the VA, health policy-makers and other stakeholders interested in improving veteran mental health outcomes.

It is worth emphasizing the high rate of detection of potentially serious psychiatric problems in this sample; we found high rates of screening positive for active PTSD (266/585, 45%), problematic drinking (243/584, 42%), major depression (164/586, 28%) and serious suicidal ideation (132/585, 23%). Facebook ads, together with other digital media advertising strategies that can support help-seeking (eg, Google AdWords) [39], may comprise critical tools in the design of effective campaigns for mental health treatment engagement. The major—and more imposing—next challenge is how to move individuals from endorsing their distress (online) to engaging in treatment in a health care or other therapeutic setting.

A critical novel component of this study was the use of an experiment, or as close to a "true experiment" as is possible in the Facebook advertising environment, to determine what ad features are most likely to result in engagement with the ad. One of the most intriguing novel findings here was that a headline encouraging users to share the ad resulted in better ad engagement. "Sharing" is a request that is uniquely suited to the social media milieu, and also appeals to a military ethos of helping peers. Results also varied depending on the level of engagement being measured and target population. What works as a "hook" regarding the generating clicks may not translate into more active participation, as was similarly found in a study reaching concerned partners of heavy drinking service members and veterans [19]. We found that for a more modest level of engagement (eg, impressions and clicks) with a broad spectrum of recent military veterans, using an image of soldiers or headlines containing a social norms message may be more

effective. In contrast, if the goal is more proactive engagement (ie, survey completion) by individuals with active suicidality, an image of a person taking a survey may work better. One reason careful development and selection of image and text may be necessary to optimize ads for individuals with suicidal ideation could be related to cognitive differences in these individuals [40]. There is, for instance, an emerging set of empirical studies showing attentional biases toward certain words [41,42].

Limitations and Future Directions

Our results should be considered in the context of several limitations. As participation in this study only involved a one-time online survey, it is not clear if the same strategies would be effective for treatment engagement [39], or engagement in research requiring a higher burden (eg, intervention or longitudinal cohort study). Also, it is possible that individuals perceived our ads in ways different than hypothesized (eg, the "survey-taking" image could have been perceived as that of "computer technology"). Military culture may also impact response to advertisements. For example, our military-related ad images favored the army, which may have contributed a higher representation of them in the sample. Future research focused on testing the effectiveness of online ads should consider a qualitative component to gain more insight into differential ad performance. If future studies can confirm and further identify ad features that result in more response and engagement by veterans with suicidal ideation, there is significant potential for targeted interventions or campaigns to enhance outreach, health messaging, help-seeking, or other behaviors.

Conclusions

Taken together, our study demonstrates that Facebook ads are an effective medium for rapidly identifying, reaching, and recruiting recent military veterans, and can particularly help in reaching individuals who screen positive for current mental health problems. These results provide a foundation to inform efforts to engage veterans disconnected from the health care system or at risk for suicidal ideation.

Acknowledgments

This project was supported by awards received by ART from the Collins Medical Trust (Research Grant) and the Medical Research Foundation of Oregon (New Investigator Grant #1603). ART is also supported by a Career Development Award from the Veterans Health Administration Health Service Research and Development (CDA 14-428) and the HSR&D Center to Improve Veteran Involvement in Care (CIVIC). The authors wish to thank Jennette Zarko and Kathryn Peck at OHSU for their assistance with the Facebook ads.

Disclaimer: The US Department of Veterans Affairs, Collins Medical Trust Research Grant, and the Medical Research Foundation of Oregon New Investigator Grant had no role in the design and conduct of the study: collection, management, analysis, and interpretation of the data, preparation, review, or approval of the manuscript, or decision to submit the manuscript for publication. The findings and conclusions in this document are those of the authors who are responsible for its contents. The findings and conclusions do not necessarily represent the views of the US Department of Veterans Affairs or the US government.

Conflicts of Interest

None declared.

References

1. Seal KH, Bertenthal D, Miner CR, Sen S, Marmar C. Bringing the war back home: mental health disorders among 103,788 US veterans returning from Iraq and Afghanistan seen at Department of Veterans Affairs facilities. Arch Intern Med 2007 Mar 12;167(5):476-482. [doi: 10.1001/archinte.167.5.476] [Medline: 17353495]
2. Reger MA, Smolenski DJ, Skopp NA, Metzger-Abamukang MJ, Kang HK, Bullman TA, et al. Risk of Suicide Among US Military Service Members Following Operation Enduring Freedom or Operation Iraqi Freedom Deployment and Separation From the US Military. JAMA Psychiatry 2015 Jun;72(6):561-569. [doi: 10.1001/jamapsychiatry.2014.3195] [Medline: 25830941]
3. Hoffmire CA, Kemp JE, Thompson C. Suicide Prevention in Patient and Nonpatient Populations: In Reply. Psychiatr Serv 2015 Oct;66(10):1120-1121. [doi: 10.1176/appi.ps.661002] [Medline: 26423166]
4. Analysis of VA Health Care Utilization among Operation Enduring Freedom (OEF), Operation Iraqi Freedom (OIF), and Operation New Dawn (OND) Veterans. 2017. (Epidemiology Program, Post-Deployment Health Group, Office of Patient Care Services, Veterans Health Administration, US Department of Veterans Affairs 2017 URL: https://www.publichealth.va.gov/docs/epidemiology/healthcare-utilization-report-fy2015-qtr3.pdf [accessed 2018-02-08] [WebCite Cache ID 6x5AC0F7g]
5. Denneson L, Corson K, Helmer D, Bair M, Dobscha S. Mental health utilization of new-to-care Iraq and Afghanistan Veterans following suicidal ideation assessment. Psych Research 2014;217:47-53. [doi: 10.1016/j.psychres.2014.03.017] [Medline: 24726814]

6. Gorman LA, Blow AJ, Ames BD, Reed PL. National Guard families after combat: mental health, use of mental health services, and perceived treatment barriers. Psychiatr Serv 2011 Jan;62(1):28-34. [doi: 10.1176/ps.62.1.pss6201_0028] [Medline: 21209296]
7. Stecker T, Fortney J, Hamilton F, Sherbourne CD, Ajzen I. Engagement in mental health treatment among veterans returning from Iraq. Patient Prefer Adherence 2010 Mar 24;4:45-49 [FREE Full text] [Medline: 20390058]
8. Bruffaerts R, Demyttenaere K, Hwang I, Chiu W, Sampson N, Kessler RC, et al. Treatment of suicidal people around the world. Br J Psychiatry 2011 Jul;199(1):64-70 [FREE Full text] [doi: 10.1192/bjp.bp.110.084129] [Medline: 21263012]
9. Downs MF, Eisenberg D. Help seeking and treatment use among suicidal college students. J Am Coll Health 2012;60(2):104-114. [doi: 10.1080/07448481.2011.619611] [Medline: 22316407]
10. Chartrand H, Robinson J, Bolton JM. A longitudinal population-based study exploring treatment utilization and suicidal ideation and behavior in major depressive disorder. J Affect Disord 2012 Dec 10;141(2-3):237-245. [doi: 10.1016/j.jad.2012.03.040] [Medline: 22703701]
11. Hynes DM, Koelling K, Stroupe K, Arnold N, Mallin K, Sohn M, et al. Veterans' access to and use of Medicare and Veterans Affairs health care. Med Care 2007 Mar;45(3):214-223. [doi: 10.1097/01.mlr.0000244657.90074.b7] [Medline: 17304078]
12. Basham C, Denneson LM, Millet L, Shen X, Duckart J, Dobscha SK. Characteristics and VA health care utilization of U.S. Veterans who completed suicide in Oregon between 2000 and 2005. Suicide Life Threat Behav 2011 Jun;41(3):287-296. [doi: 10.1111/j.1943-278X.2011.00028.x] [Medline: 21463353]
13. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]
14. Stewart J. New York Times. Facebook Has 50 Minutes of Your Time Each Day. It Wants More URL: https://www.nytimes.com/2016/05/06/business/facebook-bends-the-rules-of-audience-engagement-to-its-advantage.html [accessed 2018-06-11] [WebCite Cache ID 6x5BF3KxN]
15. Duggan M. Pew Research Center. 2015 Jan 09. Social Media Update 2014 URL: http://www.pewinternet.org/2015/01/09/social-media-update-2014/ [accessed 2018-06-11] [WebCite Cache ID 6x5BetBjf]
16. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]
17. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook Recruitment of Young Adult Smokers for a Cessation Trial: Methods, Metrics, and Lessons Learned. Internet Interv 2014 Apr;1(2):58-64. [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
18. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]
19. Pedersen E, Osilla K, Helmuth E, Tolpadi A, Gore K. Reaching Concerned Partners of Heavy Drinking Service Members and Veterans through Facebook. Military Behavioral Health 2017 Apr 13;5(3):265-273. [doi: 10.1080/21635781.2017.1316804]
20. Platt T, Platt J, Thiel DB, Kardia SLR. Facebook Advertising Across an Engagement Spectrum: A Case Example for Public Health Communication. JMIR Public Health Surveill 2016 May 30;2(1):e27 [FREE Full text] [doi: 10.2196/publichealth.5623] [Medline: 27244774]
21. Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, et al. Strategies to address participant misrepresentation for eligibility in Web-based research. Int J Methods Psychiatr Res 2014 Mar;23(1):120-129 [FREE Full text] [doi: 10.1002/mpr.1415] [Medline: 24431134]
22. Dolan P, Hallsworth M, Halpern D, King D, Metcalfe R, Vlaev I. Influencing behaviour: The mindspace way. Journal of Economic Psychology 2012 Feb;33(1):264-277. [doi: 10.1016/j.joep.2011.10.009]
23. Andreoni J. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. The Economic Journal 1990 Jun;100(401):464. [doi: 10.2307/2234133]
24. Avey J, Luthans F, Youssef C. The Additive Value of Positive Psychological Capital in Predicting Work Attitudes and Behaviors. Journal of Management 2009 Mar 17;36(2):430-452. [doi: 10.1177/0149206308329961]
25. Boamah SA, Laschinger H. The influence of areas of worklife fit and work-life interference on burnout and turnover intentions among new graduate nurses. J Nurs Manag 2016 Mar;24(2):E164-E174. [doi: 10.1111/jonm.12318] [Medline: 26122304]
26. Cialdini R. Descriptive Social Norms as Underappreciated Sources of Social Control. Psychometrika 2007 Jun 12;72(2):263-268. [doi: 10.1007/s11336-006-1560-6]
27. Gajic A, Cameron D, Hurley J. The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey. Eur J Health Econ 2012 Dec;13(6):789-799. [doi: 10.1007/s10198-011-0332-0] [Medline: 21691841]
28. Oeldorf-Hirsch A, Sundar S. Posting, commenting, and tagging: Effects of sharing news stories on Facebook. Computers in Human Behavior 2015 Mar;44:240-249. [doi: 10.1016/j.chb.2014.11.024]
29. Appendix A: National Survey of Veterans Questionnaire Instruments. URL: https://va.gov/vetdata/docs/SurveysAndStudies/AppendixAQuestionnaires.pdf [accessed 2018-06-11] [WebCite Cache ID 6x571mtX4]

30. Greenwood S, Perrin A, Duggan M. Pew Research Center. 2016 Nov 11. Social Media Update 2016 URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [accessed 2018-06-11] [WebCite Cache ID 6x5DIBQzk]
31. Prins A, Bovin MJ, Smolenski DJ, Marx BP, Kimerling R, Jenkins-Guarnieri MA, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample. J Gen Intern Med 2016 Oct;31(10):1206-1211 [FREE Full text] [doi: 10.1007/s11606-016-3703-5] [Medline: 27170304]
32. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [Medline: 9738608]
33. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
34. Joiner TE, Pfaff JJ, Acres JG. A brief screening tool for suicidal symptoms in adolescents and young adults in general health settings: reliability and validity data from the Australian National General Practice Youth Suicide Prevention Project. Behav Res Ther 2002 Apr;40(4):471-481. [Medline: 12008659]
35. von GM, Teismann T, Prinz S, Gebauer JE, Hirschfeld G. Depressive Symptom Inventory Suicidality Subscale: Optimal Cut Points for Clinical and Non-Clinical Samples. Clin Psychol Psychother 2016 Nov;23(6):543-549. [doi: 10.1002/cpp.2007] [Medline: 26857479]
36. Parker K, Cilluffo A, Stepler R. Pew Research Center. 2017. 6 facts about the U.S. military and its changing demographics URL: http://www.pewresearch.org/fact-tank/2017/04/13/6-facts-about-the-u-s-military-and-its-changing-demographics/ [accessed 2018-04-25] [WebCite Cache ID 6ywePLCDt]
37. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res 2016 Nov 7;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
38. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
39. Birnbaum ML, Garrett C, Baumel A, Scovel M, Rizvi AF, Muscat W, et al. Using Digital Media Advertising in Early Psychosis Intervention. Psychiatr Serv 2017 Nov 01;68(11):1144-1149. [doi: 10.1176/appi.ps.201600571] [Medline: 28712355]
40. Yiend J. The effects of emotion on attention: A review of attentional processing of emotional information. Cognition & Emotion 2010 Jan;24(1):3-47. [doi: 10.1080/02699930903205698]
41. Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR. Measuring the suicidal mind: implicit cognition predicts suicidal behavior. Psychol Sci 2010 Apr;21(4):511-517 [FREE Full text] [doi: 10.1177/0956797610364762] [Medline: 20424092]
42. Cha CB, Najmi S, Park JM, Finn CT, Nock MK. Attentional bias toward suicide-related stimuli predicts suicidal behavior. J Abnorm Psychol 2010 Aug;119(3):616-622 [FREE Full text] [doi: 10.1037/a0019710] [Medline: 20677851]

Abbreviations

Ad: advertisements
AUDIT-C: Test Alcohol Consumption Questions
CDA: Career Development Award
CTR: click-through rate
CIVIC: Center to Improve Veteran Involvement in Care
DSI-SS: Depressive Symptom Inventory Suicidality Subscale
OEF-OIF: Operation Enduring Freedom-Operation Iraqi Freedom
OHSU: Oregon Health & Science University
PC-PTSD: Primary Care PTSD Screen for DSM-5
PHQ-2: Patient Health Questionnaire-2
PTSD: posttraumatic stress disorder
US: United States
VA: Veterans Affairs

Edited by G Eysenbach; submitted 14.02.18; peer-reviewed by E Pedersen, RJ Westphal, J Tsai; comments to author 12.04.18; revised version received 25.04.18; accepted 13.05.18; published 05.07.18.

Please cite as:
Teo AR, Liebow SBL, Chan B, Dobscha SK, Graham AL
Reaching Those At Risk for Psychiatric Disorders and Suicidal Ideation: Facebook Advertisements to Recruit Military Veterans
JMIR Ment Health 2018;5(3):e10078
URL: http://mental.jmir.org/2018/3/e10078/
doi:10.2196/10078
PMID:29980498

Original Paper

Adapting Coordinated Anxiety Learning and Management for Veterans Affairs Community-Based Outpatient Clinics: Iterative Approach

Traci H Abraham[1,2,3], PhD; Kathy Marchant-Miros[1], BSN; Michael B McCarther[2], BA; Michelle G Craske[4,5], PhD; Geoffrey M Curran[1,2,6,7], PhD; Lisa K Kearney[8,9], PhD, ABPP; Carolyn Greene[2,10], PhD; Jan A Lindsay[11,12,13], PhD; Michael A Cucciare[1,2,3], PhD

[1]South Central Mental Illness Research, Education, and Clinical Center, US Department of Veterans Affairs, Little Rock, AR, United States

[2]Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[3]Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System, North Little Rock, AR, United States

[4]Department of Psychology, University of California, Los Angles, Los Angeles, CA, United States

[5]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, United States

[6]Center for Implementation Research, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[7]Department of Pharmacy Practice, College of Pharmacy, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[8]US Department of Veterans Affairs Center for Integrated Healthcare, Buffalo, NY, United States

[9]Department of Psychiatry, The University of Texas Health Science Center, San Antonio, TX, United States

[10]Office of Mental Health and Suicide Prevention, US Department of Veterans Affairs, Washington, DC, United States

[11]Houston Veterans Affairs Health Services Research & Development Service Center for Innovations in Quality, Effectiveness, and Safety, Michael E DeBakey Veterans Affairs Medical Center, Houston, TX, United States

[12]South Central Mental Illness Research, Education, and Clinical Center, Houston, TX, United States

[13]Menninger Department of Psychiatry & Behavioral Sciences, Houston, TX, United States

Corresponding Author:
Traci H Abraham, PhD
Center for Mental Healthcare and Outcomes Research
Central Arkansas Veterans Healthcare System
2200 Fort Roots Drive
North Little Rock, AR, 72114
United States
Phone: 1 501 257 1744
Email: Traci.Abraham@va.gov

Abstract

Background: A national priority at the US Department of Veterans Affairs (VA) is to increase the availability and accessibility of evidence-based psychotherapies (EBPs) across all VA medical facilities. Yet many veterans, particularly those who use remote outpatient VA clinics, still do not receive much needed evidence-based treatment. Strategies are needed for supporting mental health providers at rural VA community-based outpatient clinics (CBOCs) as they translate their clinical training to routine practice. The Coordinated Anxiety Learning Management (CALM) program is a computer-delivered program that supports the delivery of cognitive behavioral therapy (CBT) by providers in outpatient settings to patients with depression and anxiety, including posttraumatic stress disorder.

Objective: The objectives of our study were to (1) adapt an existing computer-based program to rural VA CBOCs through feedback from key stakeholder focus groups; (2) develop a prototype of the adapted program; and (3) determine the adapted program's acceptability and feasibility. Mental health stakeholders included VA leaders (n=4) in the implementation of EBPs, VA experts (n=4) in CBT, VA CBOC mental health providers (n=8), and veterans (n=8) diagnosed with a mental health condition treated using the CALM program and receiving treatment in a VA CBOC.

Methods: An iterative approach comprising 3 waves of focus group discussions was used to develop a modified prototype of CALM. Following each wave of focus group discussions, template analysis was used to rapidly communicate stakeholder recommendations and feedback to the design team. The original program was first adapted through a process of data collection,

design modification, and product development. Next, a prototype was developed. Finally, the redesigned program was tested for acceptability and feasibility through a live demonstration.

Results: Key stakeholders suggested modifications to the original CALM program that altered its modules' appearance by incorporating veteran-centric content. These modifications likely have no impact on the integrity of the original CALM program, but have altered its content to reflect better the demographic characteristics and experiences of rural veterans. Feedback from stakeholder groups indicates that changes will help VA patients identify with the program content, potentially enhancing their treatment engagement.

Conclusions: The development model was effective for economically gathering actionable recommendations from stakeholders to adapt a computer-based program, and it can result in the development of an acceptable and feasible computer-delivered intervention. Results have implications for developing computer-based programs targeting behavior change more broadly and enhancing engagement in EBP.

(JMIR Ment Health 2018;5(3):e10277) doi:10.2196/10277

KEYWORDS

therapy; veterans; depression; anxiety disorders; posttraumatic stress disorder; PTSD

Introduction

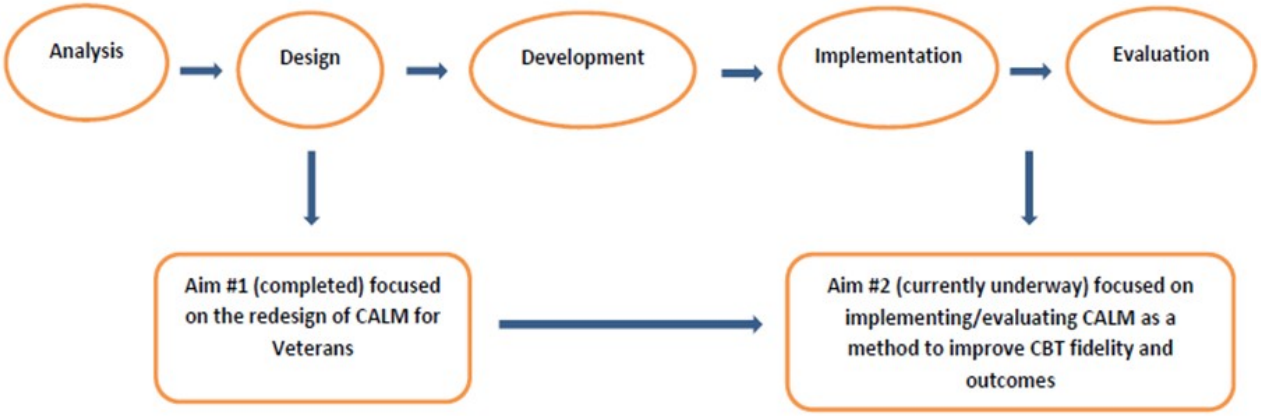
Ample evidence indicates the effectiveness of evidence-based psychotherapies (EBPs), particularly cognitive behavioral therapy (CBT) [1,2], for treating anxiety and depression, the most common mental health disorders in outpatient health care settings [3-5]. Accordingly, the US Department of Veterans Affairs (VA) has made it a national priority to increase the availability and accessibility of EBPs, particularly CBT, across all VA medical facilities and clinics to veterans needing mental health care [6]. An important component of this strategy is providing intensive, competency-based training in EBPs to VA mental health providers [7].

Despite this effort, broadly implementing CBT and other EBPs in VA treatment settings has been a challenge [8,9], especially at small or remote outpatient clinics, such as most VA community-based outpatient clinics (CBOCs). The challenges inherent in ensuring that EBPs are accessible at rural clinics are evident in results from a study published in 2010, which found that only 1 in 5 veterans with depression, anxiety, or posttraumatic stress disorder (PTSD) received at least one session of psychotherapy and that rural veterans, who are often treated in CBOCs, were even less likely than their urban counterparts to have received any psychotherapy [10]. Barriers to ensuring access to EBPs in VA CBOCs are complex, likely including practitioner factors (eg, resistance to change), training methods used, characteristics of the intervention, and organization or system factors [11]. Although evidence suggests that the disparity found in 2010 has since decreased, rural veterans still receive fewer psychotherapy sessions than their urban peers [12]. Thus, although many providers at VA CBOCs have received training in EBPs, training alone has not been sufficient to ensure that new treatments are translated into routine clinical practice [8,13].

An additional challenge at VA CBOCs is ensuring that treatments are delivered with fidelity (ie, faithfulness to the treatment model). Without adequate fidelity, patients are unlikely to experience optimal outcomes from EBPs, even where they are accessible. Although the VA has made great strides in providing EBP-specific training to providers of mental health care, training alone is not sufficient to ensure fidelity [14,15]. VA CBOCs commonly have only one mental health provider on staff. Providers are, therefore, isolated and unable to consult their peers about difficult cases. Additionally, providers may not have time to take advantage of educational resources that facilitate effective delivery of such EBPs as CBT. Providers who do not perceive themselves as proficient in delivering EBTs are sometimes reluctant to use them with patients [16]. Thus, many providers at small or remote outpatient clinics lack the resources, time, or interactions with colleagues needed to gain proficiency in these skills. Resources that can support and assist CBOC mental health providers in delivering EBPs with fidelity are needed to supplement training and ensure rural veterans' access to effective mental health treatments.

Over the last decade, the National Institute of Mental Health-funded Coordinated Anxiety Learning Management (CALM) study [17] has addressed similar challenges. The study aimed to implement CBT into non-VA outpatient clinics to support providers with little to no prior training in this treatment. To achieve this goal, researchers developed a computer-delivered program (CALM) facilitating the delivery of CBT with fidelity by mental health providers in outpatient settings [18]. CALM uses a cognitive behavioral framework including psychoeducation, cognitive restructuring, goal setting, exposure, and response prevention. It has the added advantage of being both provider and patient facing, so that patient and provider both look at the computer screen together and proceed through the modules at an individualized pace [17]. The CALM program is clinically effective for a range of anxiety disorders, including panic disorder (PD), generalized anxiety disorder (GAD), and social anxiety disorder (SAD), as well as PTSD and depression [17-19]. It may have the added benefit of helping providers maintain fidelity to the CBT treatment model [20]. Thus, implementation of the CALM program in rural VA outpatient settings may help increase veterans' receipt of efficacious CBT.

Figure 1. Modified ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model and study aims. CALM: Coordinated Anxiety Learning Management, CBT: cognitive behavioral therapy.



In this study, we sought to adapt the original CALM program for use in VA outpatient settings, thus, supporting VA CBOC mental health providers in delivering CBT to veterans with anxiety, PTSD, and depression. To achieve this objective, we modified a method common to the field of instructional design and technology (IDT) to guide the redesign of the CALM computer program for use in VA CBOCs. IDT models are prescriptive models that describe a set of activities involved in the planning, implementation, and evaluation of instructional programs [21]. Most IDT models share the core elements of analysis, design, development, implementation, and evaluation, applied in an iterative process [22,23]. The Analysis, Design, Development, Implementation, Evaluation model, as seen in Figure 1, an umbrella term referring to these common elements, formed the conceptual basis for the process used to modify the original CALM program in this study [24].

As part of this process, 3 waves of focus group discussions were conducted with 4 groups of key stakeholders to redesign and test the CALM program iteratively for use in VA CBOCs. Stakeholder groups included (1) veterans receiving recent mental health care in a VA CBOC and diagnosed with anxiety disorders and depression, (2) CBOC mental health providers, (3) expert CBT VA clinicians, and (4) VA Central Office leaders with expertise in the implementation of EBPs within VA. In this paper, we describe the iterative process used to redesign and test the modified CALM program, as well as modifications used to adapt the program for use at VA CBOCs.

Methods

Recruitment

We recruited participants from 4 key stakeholder groups: veterans (n=11) with a recent (≥1 visit in the prior 6 months) mental health care visit at a VA CBOC and with a diagnosis of one or more of the following: PD, GAD, SAD, PTSD or depression; CBOC mental health providers (n=11) in the southern region of the United States, including nurses, psychiatrists, social workers, and psychologists; VA expert CBT clinicians (n=6); and VA Central Office leaders (n=5) with expertise in implementing EBPs within VA. We conducted 3 waves of focus group discussions with each stakeholder group. Veteran participants were recruited from one CBOC in Arkansas. Mental health providers were recruited using a general call distributed via email to providers at CBOCs in Arkansas, Louisiana, Texas, and Mississippi. Multiple CBOCs were needed because most CBOCs have only one mental health provider on site. CBT experts were recruited nationally from within the VA, and leaders in implementation of EBPs within VA were recruited from the VA Central Office in Washington, DC CBT experts and leaders in implementation were individually recruited through convenience sampling based on their level of expertise and availability [25].

The purpose of recruiting veteran stakeholders was to help ensure that the modified CALM content was acceptable to the target patient population and reflected veterans' illness experiences. We included CBOC mental health providers to help ensure that the modified content was acceptable for this group of providers (who serve a largely rural veteran population), images and case studies were appropriate for the target patient population, and the navigation and flow of CALM treatment material met the needs of CBOC providers. Inclusion of VA expert CBT clinicians helped ensure that the empirical support underlying the content of the CALM program was not compromised during the modification process. Finally, inclusion of VA Central Office leaders with expertise in EBP implementation within VA helped the study team ensure that CALM was consistent with prior CBT training efforts within VA.

Data Collection

Each group of key stakeholders participated in 3 waves of separate focus group discussions. Focus group discussions are a standard method used to garner diverse feedback efficiently on novel and existing products for new product development. CBOC mental health providers, CBT experts, and VA Central Office leaders with expertise in EBP implementation within VA participated in focus group discussions via teleconference with the assistance of Lync, a Web-based meeting portal that allows participants to view the same material simultaneously. In addition to providing feedback via telephone, participants could leave written feedback on the website chat room and email

content for the focus group facilitators. Veteran focus groups met in person at a VA CBOC in Hot Springs, Arkansas.

All focus group discussions were co-moderated by the principal investigator (MAC) and a co-investigator (THA). Both researchers have extensive experience in qualitative interviewing and facilitating focus group discussions. An interview guide developed for the study was used to ensure that the discussion remained relatively consistent across stakeholder groups and that groups addressed all relevant topics. The same stakeholders were invited to participate in all 3 waves of focus group discussions to maximize the relevance of feedback (ie, participants knew whether recommendations had been correctly incorporated into the program). Focus group discussions were audiorecorded with permission from participants. The study was approved by the Central Arkansas Veterans Healthcare System Human Research Protections Program.

Data Analysis

Rapid analytic techniques informed by Sobo et al [26] and Hamilton [27] were used to quickly produce recommendations for modifying the CALM program and economically communicating modifications to the design team. Qualitative data analysts (THA and MAC) first collaborated to develop a prototype summary template in a Word document with three broad domains related to goals of the study: (1) recommendations from focus group participants; (2) evaluative observations or initial reactions or concerns, and (3) questions. They then created categories within each broad domain, reflecting various aspects of the CALM program that would be queried during focus group discussions. A copy of this prototype was then created in a new Word document for each of the 4 focus groups.

To analyze focus group discussions, the lead analyst (THA) listened to the audio recording of the first focus group discussion and systematically populated template categories with data. These data consisted of paraphrased content reflecting stakeholder recommendations, observations, reactions, concerns, and questions. The goal was to capture the full range of responses to questions and comments from focus group discussions. After the lead analyst completed summarizing all content from the focus group discussions in the template, she met with the second analyst (MAC) to discuss analytic findings. Discrepancies in how template content was summarized or categorized were resolved through discussion. This analytic process was repeated for each focus group discussion. For focus group discussions held via teleconference, analysts also incorporated written feedback from the website and emails to facilitators into the template. After analyses were finalized for all 4 focus group discussions, findings were compiled and summarized. Analysts each reviewed the summary to ensure the validity of findings. This analytic process was followed for all 3 waves of focus group discussions.

Finalizing Recommendations

A panel of experts reviewed the summary template before it was submitted to the design and development team, including study team members with expertise in CBT, the relevant mental health diagnosis, and software development. The purpose of this panel was to prioritize and determine the feasibility of each recommendation (eg, cost and time needed to complete the modification). The summary template was subsequently reduced to actionable recommendations for modification.

Description of the Original Coordinated Anxiety Learning Management Program

The CALM computerized program was created to guide and train mental health providers in delivering a course of CBT [17]. CALM was not intended, like some other technology-based interventions, to be a self-help intervention (ie, patient facing without provider involvement). Instead, provider and patient use it together, synchronously (ie, it is provider and patient facing). Provider-supported treatments such as CALM have tended to yield enhanced results compared with self-help interventions [28]. A unique feature of the CALM program is that it can be used to treat different anxiety disorders, as well as PTSD and depression. Reductions in symptoms across these conditions are accomplished through the use of basic CBT modules, which are employed across these disorders, coupled with branching modules that are disorder specific [18-20].

Results

Wave 1: Wireframe Development

During the first wave of focus group discussions, an overview of CALM was presented to each stakeholder group. This also included screenshots from the program, as well as a detailed description of the functioning of CALM for each disorder. Following the overview, stakeholders were given time to ask questions about CALM; then, they were asked to provide initial reactions or recommendations for modifying and adapting the program for use within VA CBOCs.

Summary of Wave 1 Stakeholder Feedback

Recommended modifications to the original program were largely “look and feel” changes pertaining to the images and illustrations in the modules. Stakeholders generally reported that the images and illustrations contained in the original program should be more representative of veterans. They recommended reducing the number of images and illustrations depicting college-aged women and men in white-collar occupations (eg, wearing suits) as well as incorporating imagery better reflecting the gender, age, and economic and ethnic or racial diversity of veteran patients. Veteran stakeholders did not like the background color of the CALM program and overall website template; they wanted a new logo developed specifically for the modified VA program.

Stakeholders generally also recommended modifying the images in the original program to make it more “culturally congruent” because the original program was not designed to acknowledge veterans’ military service. One expert CBT clinician noted that this can be “very invalidating” for some patients. Stakeholders suggested replacing some existing images with images of people in uniform; however, a veteran stakeholder cautioned that while incorporating images of people in uniform, “We don’t need to see a Vet on every single picture we look at.” Stakeholders also suggested using images of individuals with physical limitations in the Behavioral Activation module to encourage veterans with

prosthetics or other physical limitations to engage in physical activities.

Furthermore, stakeholders recommended revisions designed to enhance the degree to which veterans could identify with the program content. To achieve this, they suggested incorporating videos into the modules that speak to the unique mental health concerns of veterans. Stakeholders also suggested using existing VA resources, such as the National Center for PTSD's "About Face" and the VA's "Make the Connection" websites [28,29]. These websites contain videos in which veterans, family members of veterans, and mental health providers share their personal experiences with mental health concerns and relate individual stories about seeking and receiving help for mental health concerns. They also provide instructions for initiating treatment or seeking immediate help during crises. One CBT expert recommended selecting videos in which multiple veterans shared their experiences to increase the chance that patients will relate to someone in the video. A CBOC mental health provider thought that videos in which veterans describe "service or readjustment issues," such as reintegration into civilian life, would be particularly useful in helping veterans "connect" with the program content. Stakeholders also recommended incorporating links to existing Web-based psychoeducational materials developed for veterans. Suggested resources included information available at the VA National Center for PTSD and the VA websites. An additional suggestion was to include the telephone number for the national Veterans' Crisis Line at the end of the Depression Education module.

Finally, stakeholders were generally concerned that some slides seemed too "content heavy" to be engaging. Several expert CBT clinicians and CBOC mental health providers thought that having to read a large amount of material during an hour-long therapy session might be difficult, and expert CBT clinicians were concerned that younger veterans with traumatic brain injuries would be overwhelmed by busy slides. This concern was validated by feedback from veteran stakeholders, who not only recommended reducing the number of words, but removing excessive images and illustrations. One veteran noted: "Pictures: unless it means something, what's the point?"

Wave 1 Modifications to the Coordinated Anxiety Learning Management

A summary of stakeholder recommendations was presented to the expert panel for review, and actionable modifications were identified and prioritized. The study team then collaborated with the software development group to identify images and illustrations better reflecting the gender, age, and economic and ethnic or racial diversity of veterans. They also identified appropriate images of people in military uniform. The images below are examples of CALM module content before (Figure 2) and after modification (Figure 3). The alterations illustrated make use of updated imagery and of veteran stakeholder preferences for color (ie, blue rather than green) and depict an individual in uniform to enhance content relatability.

The team identified already available Web-based resources with videos of veterans describing their mental health concerns, their experiences with a condition and/or seeking mental health care, and ways that treatment helped them (Figure 4) [29]. This was realistic in the timeframe of the study, which did not allow sufficient time to develop original video content.

Finally, the team followed stakeholder suggestions in identifying links to Web-based resources for veterans and reduced the amount of text on slides identified by stakeholders as too content heavy. These modifications to the original program were submitted to the software development team, which then developed mock-ups, or wireframes, of the initial modifications to show stakeholders in Wave 2.

Wave 2: Prototype Development

The wireframes were presented during a second wave of focus group discussions with veterans (n=6), CBOC mental health providers (n=10), expert CBT clinicians (n=5), and VA leaders with expertise in the implementation of EBPs within VA (n=5). Wave 2 participants were reminded of the initial version of CALM using graphics and were provided a side-by-side comparison of all modifications. After reviewing the wireframes, stakeholders provided feedback regarding modifications from the first round of focus groups, and a few suggested additional changes to the CALM program.

Summary of Wave 2 Stakeholder Feedback

Stakeholders generally agreed that the modified program appeared more "inclusive of veterans that will be using CALM" and "user-friendly" for the veteran population. Although agreeing that the initial modifications were an improvement, one veteran stakeholder requested more images of older males to be more inclusive of Vietnam-era veterans. Stakeholders, particularly veterans, also responded positively to reductions in the amount of text on many slides. One CBOC mental health provider thought that reducing the amount of text would "leave more room for interaction" between providers and veteran patients. Stakeholders also responded positively to the inclusion of videos and links to resources. One younger participant in the veteran focus group who had completed VA treatment for PTSD suggested replacing a video in which veterans described only their treatment experiences with the one in which they explain that treatment is difficult at first, but helps. Finally, one CBOC mental health provider noted that many of his veteran patients were uncomfortable with writing and suggested the program allow veterans the option of verbally describing and recording trauma memories for use in the Exposure module, in addition to writing about them.

Wave 2 Modifications to Coordinated Anxiety Learning Management

The expert panel again reviewed a summary of recommendations and prioritized actionable modifications from Wave 2 focus group discussions. Following this, the study team identified images and videos per stakeholder recommendations. The principal investigator (MAC) collaborated with one CALM developer (MAC) to incorporate an option allowing veterans who are uncomfortable with writing to verbally describe and record their traumatic experiences (Figure 5). These modifications were submitted to the design team, and a prototype of the modified CALM program was developed.

Figure 2. Original Coordinated Anxiety Learning Management (CALM) module content. (Screenshot taken by KMM; Source: University of California, Los Angeles).

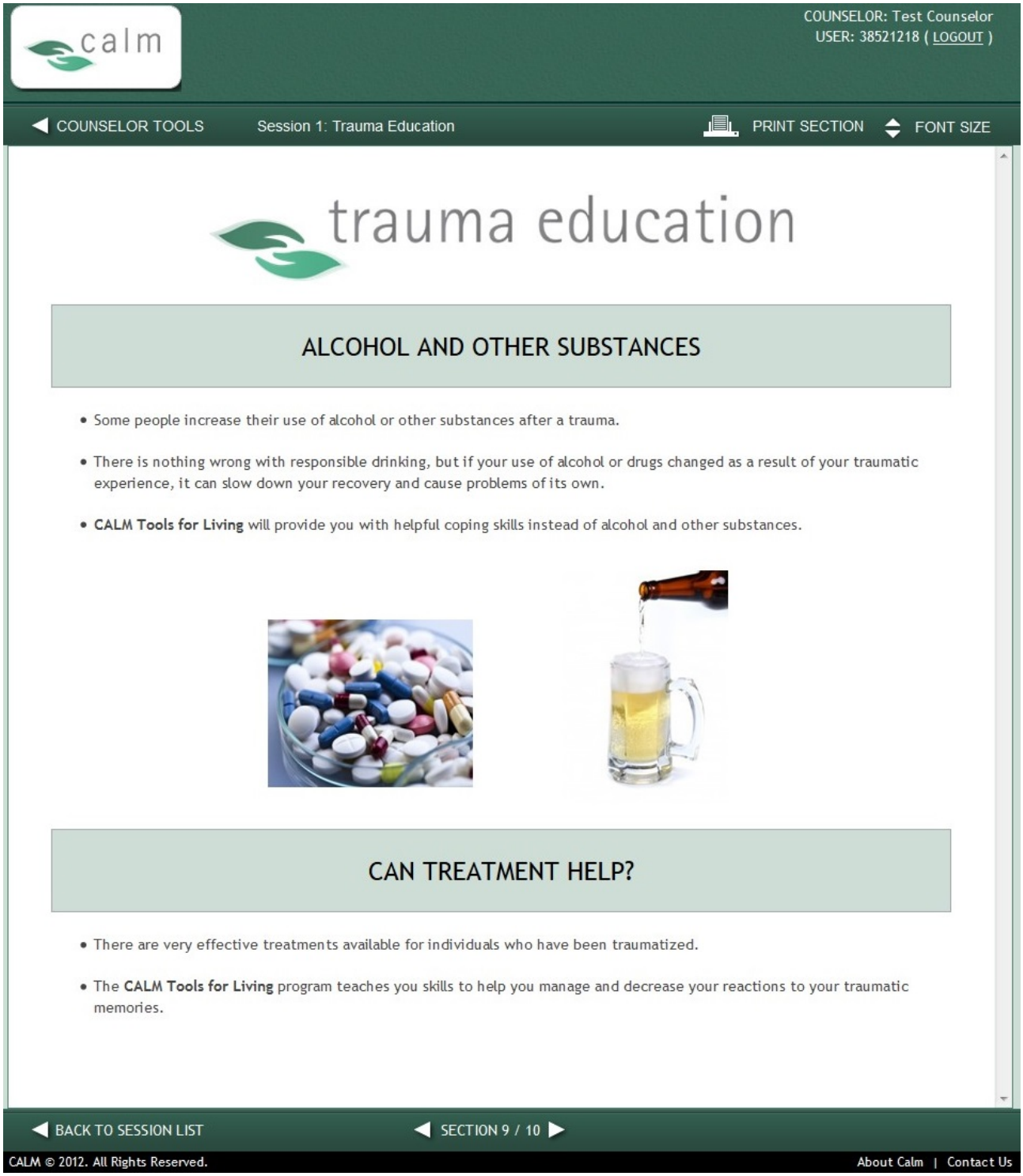


Figure 3. Modified Coordinated Anxiety Learning Management (CALM) module content. (Screenshot taken by KMM; Source: US Department of Veterans Affairs).

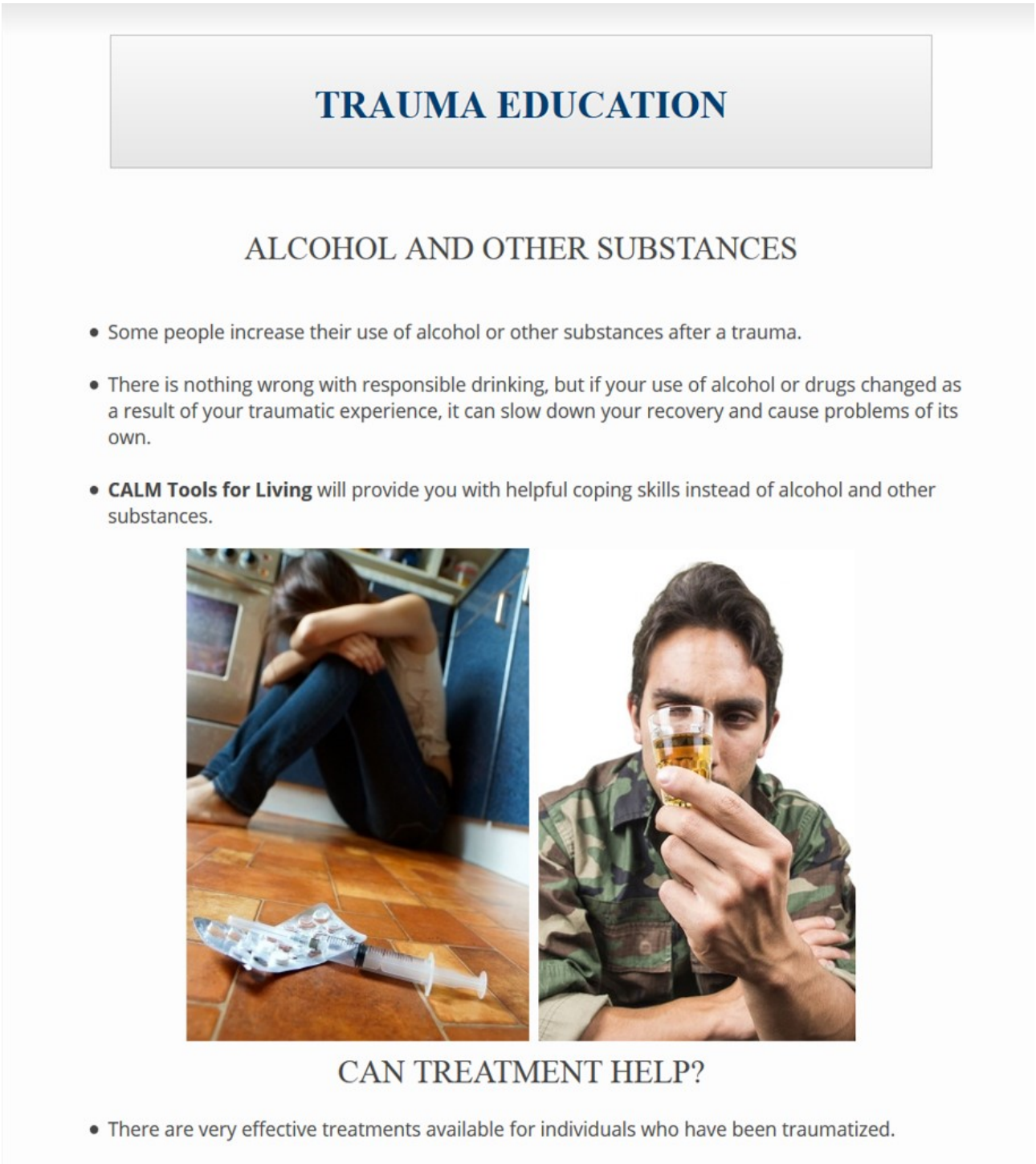


Figure 4. Video featuring veterans describing their experiences with posttraumatic stress disorder (PTSD); used with permission from the National Center for PTSD "About Face" website. (Screenshot taken by KMM; Source: US Department of Veterans Affairs).

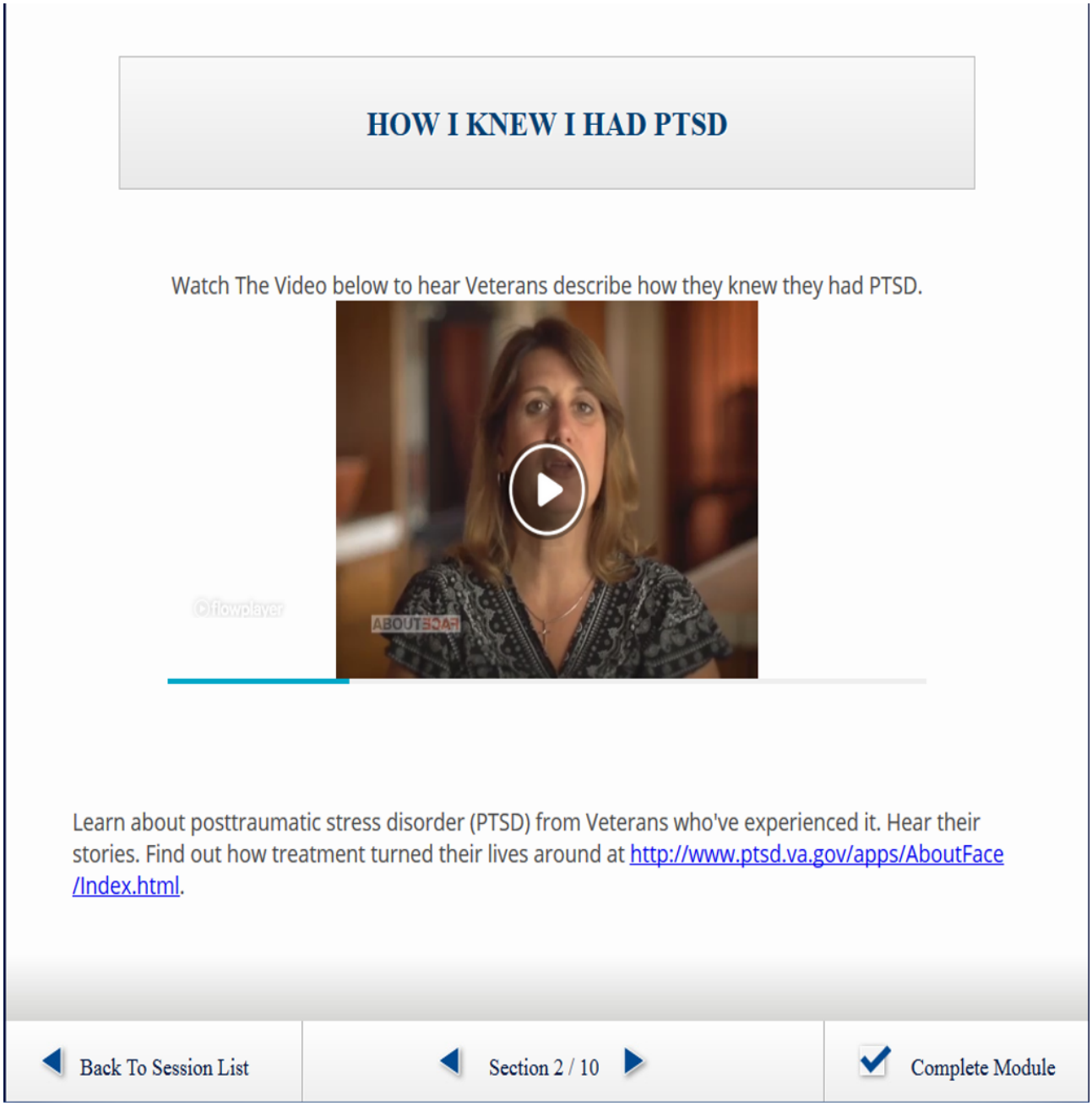
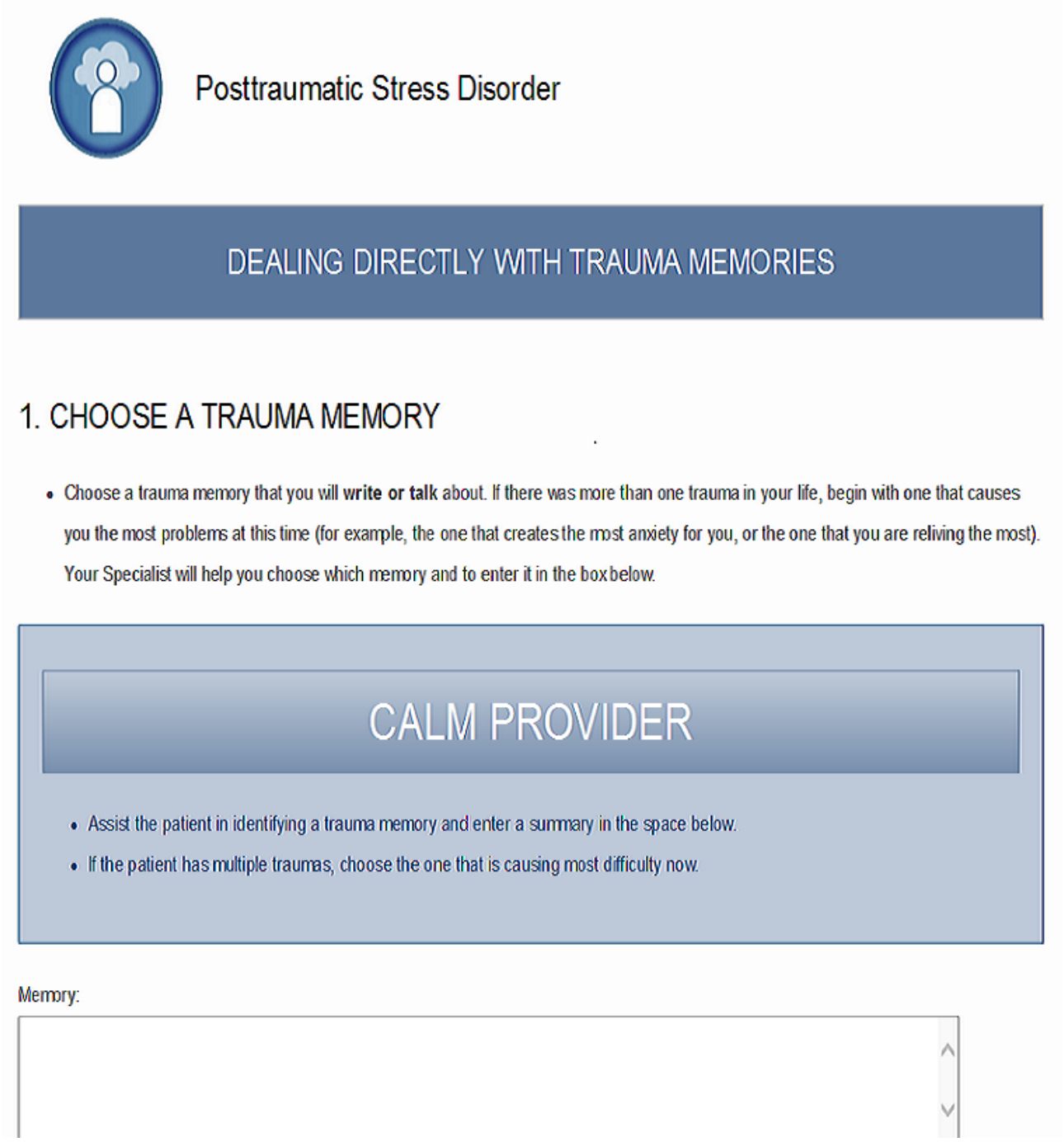


Figure 5. Option allowing veterans to describe and record traumatic experiences. (Screenshot taken by KMM; Source: US Department of Veterans Affairs).



Wave 3: Validation of Coordinated Anxiety Learning Management Modifications

During a final wave of focus group discussions, the prototype was pilot-tested by MGC during a live teleconferenced demonstration to 3 stakeholder groups: CBOC mental health providers (n=6), expert CBT clinicians (n=4), and VA leaders in clinic operations and implementation (n=3). Minor problems that arose during the pilot test were recorded in written notes by the study team and rapidly communicated to the design team following the demonstration. Veteran stakeholders (n=4) reviewed modified CALM modules presented in person by the principal investigator (MAC) and looked at an overview of all modifications by the lead qualitative analyst (THA). No remaining concerns about the program were elicited through this final wave of focus groups.

Discussion

Principal Findings

In this study, we used an iterative approach to adapt the original CALM program for use within the VA and, particularly, at rural, CBOCs. Overall, key stakeholders suggested modifications to the original CALM program that were largely “look and feel” adaptations that altered the appearance of the modules by incorporating veteran-centric content. This included relatively

simple changes, such as replacing images and developing a new template with the VA logo, as well as more significant adaptations, such as embedding videos of veterans describing their treatment and illness experiences and modifying the case studies to better reflect experiences common to rural veterans. These modifications likely have no impact on the integrity of the original CALM program, while altering the content to better reflect the demographic characteristics and experiences of rural veterans.

An additional adaptation was to customize the treatment content to allow veterans the option of orally recounting (as opposed to only writing) their trauma experiences. This modification, along with the overall flexibility of the CALM program to meet the treatment needs of veterans with a wide variety of mental health conditions, is consistent with a long-standing cultural shift at the VA toward the provision of patient-centered health care consisting of a menu of treatment options that can be tailored to each individual veteran's needs and goals for health and well-being [30,31]. Patient-centered care can improve health outcomes and increase patient satisfaction and self-management of chronic conditions [32,33]. It has also been associated with decreased health care utilization, including the annual number of specialty care visits, less frequent hospitalizations, and fewer laboratory and diagnostic tests [34].

Feedback from stakeholder groups—and, in particular, rural veterans—indicates that these modifications will help VA patients identify with the program content. This could potentially improve patient-provider communication during therapy sessions and increase veteran engagement in EBP for anxiety, depression, and PTSD. Enhancing engagement is important because attrition from EBPs is high among veterans, particularly those with PTSD [35]. In one recent study, of 351 veterans who initiated EBP for PTSD, one third (n=135) dropped out before treatment completion [36]. Because completion of EBP not only significantly reduces mental health symptoms [37] but also positively impacts physical health and functioning while decreasing PTSD-related health care costs [38,39], enhancing veteran engagement is critical. Indeed, in one study, the original CALM program was not only found to be acceptable to both providers and patients but also resulted in substantial treatment engagement and homework compliance [20].

CALM has the added benefit of being both patient and provider facing; thus, providers—not patients—directly interact with the program. The use of CALM during treatment sessions does not require patients to possess either technological proficiency or access to the internet at home. Future research could explore whether the program helps older patients gain a sense of familiarity with the use of technology in health care, thus, reducing the potential for a deepening digital divide among older adults.

Feedback, comments, and suggestions obtained from the 4 stakeholder groups, in tandem with results from similar studies [40-44], suggest a few general principles for developing or enhancing the acceptability of technology-based interventions in general. Computer- and Web-based programs should use text judiciously to reduce the potential for boredom and fatigue from reading large amounts of information [40,41]. Additionally, images and illustrations should be up-to-date [40,41], and health information should be tailored to the target audience and individualized to the patients to maximize the effect of technology-based interventions on behavior change [42,43]. Mental health interventions that receive support from providers enhance patients' willingness to initiate computer- and internet-based treatments [44]. Because CALM is both provider and patient facing, provider support for the intervention is implicit. Patient-facing programs, however, will likely need providers to encourage their patients to initiate technology-delivered EBPs. Finally, adaptations should be based on feedback from key stakeholders, and program modifications should be presented to and reviewed by the same stakeholders.

The study design draws upon numerous strengths. An iterative design that included inviting the same stakeholders to participate in 3 waves of focus group discussions helped ensure the relevance and consistency of recommended modifications. The strength of this approach was reinforced by hosting the veteran focus group last, which allowed veterans to respond to the recommendations from the other stakeholder groups. Additionally, collecting feedback from stakeholders who are familiar with rural veterans, EBPs, and clinical practice at VA CBOCs may have increased the acceptability and feasibility of the modified program for use in this population. This approach was aligned with a person-based approach for developing and tailoring technology-based interventions [45]. Finally, using a rapid analytic technique allowed the study team to economically communicate stakeholder recommendations to the design team, ensuring that modifications were made within project time constraints.

Limitations

One limitation of our study was that focus group participation declined throughout the study, particularly among veteran stakeholders. Because we were unable to obtain feedback from every stakeholder who participated in the first wave of focus group discussions, it is possible that we may have omitted or failed to implement suggestions as originally envisioned. Additionally, veteran and mental health provider stakeholders were recruited from one geographical region in the southern United States (Arkansas, Louisiana, and Texas). Thus, stakeholder feedback may not be generalizable to other locations. Although a necessary first step in this direction, this study also does not provide data regarding fidelity to the CBT model or treatment outcomes. A next step in this line of research is to determine whether the modified version of CALM improves VA CBOC mental health providers' fidelity to the CBT model and improves veteran outcomes. The evidence already indicates that the original CALM program achieves these objectives [18,20]; thus, it is important to ascertain whether this has been maintained (or perhaps even enhanced) following revisions. It will also be important to assess whether implementing the program helps ensure the translation of CALM into routine clinical practice in the future because it may increase veterans' access to EBPs at VA CBOCs.

Conclusions

We modified the CALM program for use in rural VA clinics based on feedback from 2 waves of focus group discussions

with 4 key stakeholder groups. The results of pilot testing the modified program during a third and final wave suggest that the adaptations increased the relevance and acceptability of CALM content for rural veterans and other key VA stakeholders, such as mental health providers at VA CBOCs. It will be important to assess whether using the program as an interface between providers and patients during sessions enhances veterans' engagement in EBPs. The iterative approach used in this study holds promise for economically gathering actionable recommendations from stakeholders to enhance the acceptability and feasibility of computer-based programs in health care settings.

Acknowledgments

This work was supported by funding from the US Department of Veterans Affairs Health Services Research CREATE initiative, Development Service to MAC, and the use of facilities and resources of the VA HSR&D Center for Innovations in Quality, Effectiveness and Safety (Cin13-413). The views expressed are those of the authors and do not represent the views of the US Department of Veterans Affairs, the US government, or Baylor College of Medicine.

Authors' Contributions

THA and MAC co-wrote the introduction and methods, and THA wrote the results and discussion. All authors reviewed, edited, and approved the final manuscript. THA, MAC, KMM, and MBM collected all data. All authors were involved in developing the modified program. All authors were involved in modification of CALM for use in VA CBOCs.

Conflicts of Interest

None declared.

References

1. Cape J, Whittington C, Buszewicz M, Wallace P, Underwood L. Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. BMC Med 2010 Jun 25;8:38 [FREE Full text] [doi: 10.1186/1741-7015-8-38] [Medline: 20579335]
2. Høifødt RS, Strøm C, Kolstrup N, Eisemann M, Waterloo K. Effectiveness of cognitive behavioural therapy in primary health care: a review. Fam Pract 2011 Oct;28(5):489-504 [FREE Full text] [doi: 10.1093/fampra/cmr017] [Medline: 21555339]
3. Leon AC, Olfson M, Broadhead WE, Barrett JE, Blacklow RS, Keller MB, et al. Prevalence of mental disorders in primary care. Implications for screening. Arch Fam Med 1995 Oct;4(10):857-861. [Medline: 7551133]
4. Ansseau M, Dierick M, Buntinkx F, Cnockaert P, De Smedt J, Van Den Haute M, et al. High prevalence of mental disorders in primary care. J Affect Disord 2004 Jan;78(1):49-55. [Medline: 14672796]
5. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [Medline: 17339617]
6. US Department OVA. Uniform Mental Health Services in VA Medical Centers and Clinics. In: VHA Handbook 1160.01. Washington, DC: US Department of Veterans Affairs; 2008.
7. Karlin BE, Ruzek JI, Chard KM, Eftekhari A, Monson CM, Hembree EA, et al. Dissemination of evidence-based psychological treatments for posttraumatic stress disorder in the Veterans Health Administration. J Trauma Stress 2010 Dec;23(6):663-673. [doi: 10.1002/jts.20588] [Medline: 21171126]
8. Sullivan G, Blevins D, Kauth MR. Translating clinical training into practice in complex mental health systems: toward opening the 'black box' of implementation. Implement Sci 2008 Jun 03;3:33 [FREE Full text] [doi: 10.1186/1748-5908-3-33] [Medline: 18522742]
9. Kauth MR, Sullivan G, Cully J, Blevins D. Facilitating practice changes in mental health clinics: A guide for implementation development in health care systems. Psychological Services 2011;8(1):36-47. [doi: 10.1037/a0022250]
10. Cully JA, Jameson JP, Phillips LL, Kunik ME, Fortney JC. Use of psychotherapy by rural and urban veterans. J Rural Health 2010;26(3):225-233. [doi: 10.1111/j.1748-0361.2010.00294.x] [Medline: 20633090]
11. Ruzek JI, Rosen RC. Disseminating evidence-based treatments for PTSD in organizational settings: A high priority focus area. Behav Res Ther 2009 Nov;47(11):980-989. [doi: 10.1016/j.brat.2009.07.008] [Medline: 19632668]
12. Mott JM, Grubbs KM, Sansgiry S, Fortney JC, Cully JA. Psychotherapy Utilization Among Rural and Urban Veterans From 2007 to 2010. J Rural Health 2015;31(3):235-243. [doi: 10.1111/jrh.12099] [Medline: 25471067]
13. Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2009(2):CD003030. [doi: 10.1002/14651858.CD003030.pub2] [Medline: 19370580]
14. Stetler CB, Legro MW, Wallace CM, Bowman C, Guihan M, Hagedorn H, et al. The role of formative evaluation in implementation research and the QUERI experience. J Gen Intern Med 2006 Feb;21 Suppl 2:S1-S8 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00355.x] [Medline: 16637954]

15. Jerrell JM, Ridgely MS. Impact of robustness of program implementation on outcomes of clients in dual diagnosis programs. Psychiatr Serv 1999 Jan;50(1):109-112. [doi: 10.1176/ps.50.1.109] [Medline: 9890592]
16. Foa EB, Gillihan SJ, Bryant RA. Challenges and Successes in Dissemination of Evidence-Based Treatments for Posttraumatic Stress: Lessons Learned From Prolonged Exposure Therapy for PTSD. Psychol Sci Public Interest 2013 May;14(2):65-111 [FREE Full text] [doi: 10.1177/1529100612468841] [Medline: 25722657]
17. Sullivan G, Craske MG, Sherbourne C, Edlund MJ, Rose RD, Golinelli D, et al. Design of the Coordinated Anxiety Learning and Management (CALM) study: innovations in collaborative care for anxiety disorders. Gen Hosp Psychiatry 2007;29(5):379-387 [FREE Full text] [doi: 10.1016/j.genhosppsych.2007.04.005] [Medline: 17888803]
18. Roy-Byrne P, Craske MG, Sullivan G, Rose RD, Edlund MJ, Lang AJ, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. JAMA 2010 May 19;303(19):1921-1928 [FREE Full text] [doi: 10.1001/jama.2010.608] [Medline: 20483968]
19. Craske MG, Stein MB, Sullivan G, Sherbourne C, Bystritsky A, Rose RD, et al. Disorder-specific impact of coordinated anxiety learning and management treatment for anxiety disorders in primary care. Arch Gen Psychiatry 2011 Apr;68(4):378-388 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2011.25] [Medline: 21464362]
20. Craske MG, Rose RD, Lang A, Welch SS, Campbell-Sills L, Sullivan G, et al. Computer-assisted delivery of cognitive behavioral therapy for anxiety disorders in primary-care settings. Depress Anxiety 2009 Mar;26(3):235-242 [FREE Full text] [doi: 10.1002/da.20542] [Medline: 19212970]
21. Hilgart MM, Ritterband LM, Thorndike FP, Kinzie MB. Using instructional design process to improve design and development of Internet interventions. J Med Internet Res 2012;14(3):e89 [FREE Full text] [doi: 10.2196/jmir.1890] [Medline: 22743534]
22. Reiser R, Dempsey J. Trends and issues in instructional design and technology. Upper Saddle River, NJ: Merrill/Prentice Hall; 2002:0130222976.
23. Gagne RM, Wager WW, Golas KC, Keller JM, Russell JD. Principles of instructional design, 5th edition. Nonprofit Management Leadership 2005 Feb;44(2):44-46. [doi: 10.1002/pfi.4140440211]
24. Molenda M. In search of the elusive ADDIE model. Perf. Improv 2003 May;42(5):34-36. [doi: 10.1002/pfi.4930420508]
25. Bernard H. Research methods in anthropology: Qualitative and quantitative approaches. Third Edition. Walnut Creek, CA: Alta Mira Press; 2002:A.
26. Sobo EJ, Seid M, Reyes GL. Parent-identified barriers to pediatric health care: a process-oriented model. Health Serv Res 2006 Feb;41(1):148-172 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00455.x] [Medline: 16430605]
27. Hamilton A. US Department of Veterans Affairs. Qualitative methods in rapid turn-around health services research URL: https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=780 [WebCite Cache ID 6xWSS7giR]
28. 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]
29. National CFP. About Face. URL: https://www.ptsd.va.gov/apps/AboutFace/ [accessed 2018-02-23] [WebCite Cache ID 6xRxkidLV]
30. Kuehn BM. Veterans health system cited by experts as a model for patient-centered care. JAMA 2012 Feb 01;307(5):442-443. [doi: 10.1001/jama.2012.47] [Medline: 22298663]
31. Perlin JB, Kolodner RM, Roswell RH. The Veterans Health Administration: quality, value, accountability, and information as transforming strategies for patient-centered care. Am J Manag Care 2004 Nov;10(11 Pt 2):828-836 [FREE Full text] [Medline: 15609736]
32. McMillan SS, Kendall E, Sav A, King MA, Whitty JA, Kelly F, et al. Patient-centered approaches to health care: a systematic review of randomized controlled trials. Med Care Res Rev 2013 Dec;70(6):567-596. [doi: 10.1177/1077558713496318] [Medline: 23894060]
33. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013 Aug;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]
34. Bertakis KD, Azari R. Patient-centered care is associated with decreased health care utilization. J Am Board Fam Med 2011;24(3):229-239 [FREE Full text] [doi: 10.3122/jabfm.2011.03.100170] [Medline: 21551394]
35. Mott JM, Mondragon S, Hundt NE, Beason-Smith M, Grady RH, Teng EJ. Characteristics of U.S. veterans who begin and complete prolonged exposure and cognitive processing therapy for PTSD. J Trauma Stress 2014 Jun;27(3):265-273. [doi: 10.1002/jts.21927] [Medline: 24948535]
36. Kehle-Forbes SM, Meis LA, Spoont MR, Polusny MA. Treatment initiation and dropout from prolonged exposure and cognitive processing therapy in a VA outpatient clinic. Psychol Trauma 2016 Jan;8(1):107-114. [doi: 10.1037/tra0000065] [Medline: 26121175]
37. Tuerk PW, Wangelin B, Rauch SAM, Dismuke CE, Yoder M, Myrick H, et al. Health service utilization before and after evidence-based treatment for PTSD. Psychol Serv 2013 Nov;10(4):401-409. [doi: 10.1037/a0030549] [Medline: 23148769]
38. Meyers LL, Strom TQ, Leskela J, Thuras P, Kehle-Forbes SM, Curry KT. Service utilization following participation in cognitive processing therapy or prolonged exposure therapy for post-traumatic stress disorder. Mil Med 2013 Jan;178(1):95-99. [Medline: 23356126]

39. Rauch SAM, Grunfeld TEE, Yadin E, Cahill SP, Hembree E, Foa EB. Changes in reported physical health symptoms and social function with prolonged exposure therapy for chronic posttraumatic stress disorder. Depress Anxiety 2009;26(8):732-738. [doi: 10.1002/da.20518] [Medline: 18781660]
40. Cucciare MA, Jamison AL, Combs AS, Joshi G, Cheung RC, Rongey C, et al. Adapting a computer-delivered brief alcohol intervention for veterans with Hepatitis C. Inform Health Soc Care 2017 Dec;42(4):378-392. [doi: 10.1080/17538157.2016.1255628] [Medline: 28068154]
41. Batterham PJ, Calear AL. Preferences for Internet-Based Mental Health Interventions in an Adult Online Sample: Findings From an Online Community Survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
42. Kinzie MB. Instructional design strategies for health behavior change. Patient Educ Couns 2005 Jan;56(1):3-15. [doi: 10.1016/j.pec.2004.02.005] [Medline: 15590217]
43. Morrison L, Moss-Morris R, Michie S, Yardley L. Optimizing engagement with Internet-based health behaviour change interventions: comparison of self-assessment with and without tailored feedback using a mixed methods approach. Br J Health Psychol 2014 Nov;19(4):839-855 [FREE Full text] [doi: 10.1111/bjhp.12083] [Medline: 24308806]
44. Lapham GT, Hawkins EJ, Chavez LJ, Achtmeyer CE, Williams EC, Thomas RM, et al. Feedback from recently returned veterans on an anonymous web-based brief alcohol intervention. Addict Sci Clin Pract 2012 Aug 28;7:17 [FREE Full text] [doi: 10.1186/1940-0640-7-17] [Medline: 23186354]
45. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

Abbreviations

CALM: Coordinated Anxiety Learning Management
CBOC: community-based outpatient clinics
CBT: cognitive behavioral therapy
EBP: evidence-based psychotherapies
GAD: generalized anxiety disorder
IDT: instructional design and technology
PD: panic disorder
PTSD: posttraumatic stress disorder
SAD: social anxiety disorder
VA: US Department of Veterans Affairs

Edited by J Torous; submitted 01.03.18; peer-reviewed by M Pritchard, M Subotic-Kerry; comments to author 02.05.18; revised version received 28.05.18; accepted 29.05.18; published 22.08.18.

Please cite as:
Abraham TH, Marchant-Miros K, McCarther MB, Craske MG, Curran GM, Kearney LK, Greene C, Lindsay JA, Cucciare MA
Adapting Coordinated Anxiety Learning and Management for Veterans Affairs Community-Based Outpatient Clinics: Iterative Approach
JMIR Ment Health 2018;5(3):e10277
URL: http://mental.jmir.org/2018/3/e10277/
doi:10.2196/10277
PMID:30135051

Original Paper

Comparing Internet-Based Cognitive Behavioral Therapy With Standard Care for Women With Fear of Birth: Randomized Controlled Trial

Elisabet Rondung[1*], MS Psych; Elin Ternström[2*], RN, RM, PhD; Ingegerd Hildingsson[2,3], RN, RM, PhD; Helen M Haines[4], RN, RM, BN, MPH, PhD; Örjan Sundin[1], MS Psych, PhD; Johanna Ekdahl[1], MS Psych, PhD; Annika Karlström[3], RN, RM, PhD; Birgitta Larsson[3], RN, RM, PhD; Birgitta Segeblad[5], MD, PhD; Rebecca Baylis[2], RN, RM; Christine Rubertsson[2,6], RN, RM, PhD

[1]Department of Psycholgy, Mid Sweden University, Östersund, Sweden
[2]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[3]Department of Nursing, Mid Sweden University, Sundsvall, Sweden
[4]Department of Rural Health, The University of Melbourne, Victoria, Australia
[5]Department of Obstetrics and Gynecology, Uppsala University Hospital, Uppsala, Sweden
[6]Department of Health Science, Faculty of Medicine, Lund university, Lund, Sweden
[*]these authors contributed equally

Corresponding Author:
Elisabet Rondung, MS Psych
Department of Psycholgy
Mid Sweden University
Campus Östersund
Östersund, 831 25
Sweden
Phone: 46 706430721
Fax: 46 10 142 8004
Email: elisabet.rondung@miun.se

Abstract

Background: Although many pregnant women report fear related to the approaching birth, no consensus exists on how fear of birth should be handled in clinical care.

Objective: This randomized controlled trial aimed to compare the efficacy of a guided internet-based self-help program based on cognitive behavioral therapy (guided ICBT) with standard care on the levels of fear of birth in a sample of pregnant women reporting fear of birth.

Methods: This nonblinded, multicenter randomized controlled trial with a parallel design was conducted at three study centers (hospitals) in Sweden. Recruitment commenced at the ultrasound screening examination during gestational weeks 17-20. The therapist-guided ICBT intervention was inspired by the Unified protocol for transdiagnostic treatment of emotional disorders and consisted of 8 treatment modules and 1 module for postpartum follow-up. The aim was to help participants observe and understand their fear of birth and find new ways of coping with difficult thoughts and emotions. Standard care was offered in the three different study regions. The primary outcome was self-assessed levels of fear of birth, measured using the Fear of Birth Scale.

Results: We included 258 pregnant women reporting clinically significant levels of fear of birth (guided ICBT group, 127; standard care group, 131). Of the 127 women randomized to the guided ICBT group, 103 (81%) commenced treatment, 60 (47%) moved on to the second module, and only 13 (10%) finished ≥4 modules. The levels of fear of birth did not differ between the intervention groups postintervention. At 1-year postpartum follow-up, participants in the guided ICBT group exhibited significantly lower levels of fear of birth (U=3674.00, z=−1.97, P=.049, Cohen d=0.28, 95% CI –0.01 to 0.57). Using the linear mixed models analysis, an overall decrease in the levels of fear of birth over time was found (P≤ .001), along with a significant interaction between time and intervention, showing a larger reduction in fear of birth in the guided ICBT group over time ($F_{1,192.538}$=4.96, P=.03).

Conclusions: Fear of birth decreased over time in both intervention groups; while the decrease was slightly larger in the guided ICBT group, the main effect of time alone, regardless of treatment allocation, was most evident. Poor treatment adherence to guided ICBT implies low feasibility and acceptance of this treatment.

Trial Registration: ClinicalTrials.gov NCT02306434; https://clinicaltrials.gov/ct2/show/NCT02306434 (Archived by WebCite at http://www.webcitation.org/70sj83qat)

(JMIR Ment Health 2018;5(3):e10420) doi:10.2196/10420

KEYWORDS

fear of birth; anxiety; pregnancy; cognitive behavioral therapy; internet-based

Introduction

Background

Fear of birth (FOB) has been recognized as an important component in psychosocial antenatal care. More than just affecting the emotional well-being of pregnant women, FOB has repeatedly been associated with measures of both anxiety and depression [1-3], as with obstetric complications [4], negative birth experiences [4,5], and requests for cesarean births [6,7].

Depending on the population studied and the measurement tool used for identification, the prevalence of FOB varies. However, a worldwide FOB prevalence of 14% has recently been found [8]. Commonly, primiparous women report slightly higher fear levels than multiparous women [9]. It remains unclear how levels of FOB change throughout pregnancy. Huizink [10] found that the mean levels of FOB decreased from early to midpregnancy and were elevated again in late pregnancy. On the individual level, however, different patterns, with fear levels increasing or decreasing during pregnancy, have been shown [11]. The distress experienced because of FOB can persist beyond giving birth. Women with FOB during pregnancy are at risk of still feeling that fear as long as 1 year postpartum or in a subsequent pregnancy [12,13].

Although widely acknowledged in clinical care, the concept of FOB remains poorly defined [14,15], and several terms such as fear of childbirth [16,17], tokophobia [18], or pregnancy anxiety [19] are being used. In essence, FOB refers to experiences of fear, anxiety, or worry related to giving birth. Little is known about the psychological constructs explaining FOB [9]. A distinction is commonly made between fear that predates first pregnancy—primary FOB—and fear that appears following traumatic or distressing childbirth—secondary FOB [18]. Many other variations exist among women fearing birth with regard to fear acquisition, fear objects, symptom severity, and comorbidity [20-23]. Thus, it is important that interventions for FOB are broad or adequately flexible to meet this heterogeneity.

No consensus exists on how FOB should be handled in clinical care. A few different treatment protocols have been evaluated in randomized controlled trials (RCTs). When comparing psychoeducational group sessions with standard care (SC) for women with severe FOB, fewer cesarean births, more positive birth experiences, and less depressive symptoms postpartum were found in the intervention group, however, with small effect sizes for the psychological variables [24-26]. An Australian trial compared midwife telephone counseling on two occasions to standard antenatal care, showing reduced levels of FOB at postintervention (gestational week 36) in both groups [27]. After adjustment for the preintervention levels of FOB, the reduction in FOB was slightly higher for women in the intervention group. In Sweden, women with FOB during pregnancy are, after referral, offered counseling by specially trained midwives [28]. Larsson et al [5] reported that 1 year postpartum most women who received this counseling for FOB were satisfied with their care; however, their pregnancies and births were less favorable than nonfearful women in terms of their level of fear, degree of positive birth experiences, and rate of elective cesarean births. Although generally showing some positive effects of interventions targeting FOB, none of these studies has been convincing in reducing FOB. Given the apparent associations between FOB and measures of anxiety and depression, treatment protocols known to be efficacious in reducing fear, anxiety, and depressive symptoms are thus important to explore.

To date, cognitive behavioral therapy (CBT) remains the treatment of choice for most anxiety disorders [29-33] and one of the treatment alternatives recommended for depression [34]. These recommendations apply to women in the antenatal and postpartum period [35].

Recent advances in the field of CBT offer two highly interesting treatment alternatives with regard to FOB. First, transdiagnostic CBT treatment protocols have shown to be as efficacious as diagnosis-specific interventions for anxiety disorders, with robust effects even in the presence of comorbidity [36]. Considering the various associations between FOB and measures of anxiety and depression, the lack of knowledge regarding specific psychological mechanisms underpinning FOB and the apparent heterogeneity with regard to the symptom severity, comorbidity, and anxious focus, a transdiagnostic approach to CBT might be especially suitable in this context. Second, while the evidence is not yet conclusive [37], evaluations of interventions building on the principles and techniques of CBT but provided over the internet suggest equivalency with face-to-face CBT in terms of efficacy [38]. Guided internet-based self-help programs are well accepted by patients [39] and can be advantageous with regard to patients' access to treatment, the amount of therapist time required, and their cost-effectiveness [40]. These advantages could be important when trying to implement a new treatment approach in psychosocial antenatal care. Additionally, a treatment that is flexible with regard to time and location might well suit the needs of expecting mothers and families.

Although internet-based self-help based on the principles of CBT could hold promise as a treatment alternative for women experiencing FOB, only one earlier study investigating the feasibility of such an approach has been published. In this nonrandomized study, Nieminen et al [41] tested an internet-based CBT self-help program for primiparous women with FOB. The authors reported a within-group decrease in FOB from preintervention to postintervention (Cohen *d*=0.95) and concluded that internet-based cognitive behavioral therapy (ICBT) has potential in the treatment of FOB for motivated primiparous women; however, they recommended confirmation by randomized studies.

Objective

The primary aim of this RCT was to evaluate the efficacy of a guided internet-based self-help program based on CBT compared with SC on the levels of FOB in late pregnancy and 1 year postpartum in a Swedish sample of primiparous and multiparous women reporting clinically significant levels of FOB.

Methods

Design and Setting

This RCT was associated with the Uppsala University Psychosocial Care Program (U-CARE). This study, called the U-CARE Pregnancy trial, was a nonblinded, multicenter RCT with a parallel design, comparing guided ICBT with SC for pregnant women reporting FOB [42]. It was registered at ClinicalTrials.gov (NCT02306434) and approved by the Regional Ethical Review Board in Uppsala (No. 2013/209). We used a study-specific website called the U-CARE portal [43] for data collection and implementation of the guided ICBT intervention. Once the study was launched, the methods used for data collection and internet-based intervention were frozen and could not be changed. Recruitment and SC interventions were conducted at three study centers in Sweden—1 university hospital with an annual rate of 4000 births and 2 referral hospitals with an annual rate of 2800 and 1600 births, respectively.

Sample Size Estimation

The sample size was determined on a reduction in the level of FOB, assessed in midpregnancy and 1 year after giving birth. The sample size of this study was based on a Swedish study, where 59% of women who had FOB during pregnancy reported no FOB 1 year postpartum [44]. With a 20% reduction of FOB, a two-sided test, a power of.80, and a significance level of 5%, the power calculation showed that approximately 200 participants needed to be enrolled in this study [42].

Participants

Between February 2014 and February 2015, women undergoing ultrasound screening examination in gestational weeks 17-20 were screened for possible identification of FOB. The level of FOB was assessed by the Fear of Birth Scale (FOBS), where a cutoff of ≥60 was used to identify FOB [11,45]. The inclusion criteria were an ongoing pregnancy in gestational weeks 17-20, an ultrasound screening examination with no reported adverse findings, FOBS≥60, proficiency in Swedish language, and personal access to a mobile phone and computer with internet connection. Before enrollment, eligible women were given written and oral information about the study by the research midwives, and women willing to participate gave their written informed consent. Those who gave their consent received log-in details to the U-CARE portal and logged in and completed the internet-based preintervention questionnaire. After the completion of the questionnaire, participants were randomized by the U-CARE portal (1:1) to either the guided ICBT or the SC group.

Of 4502 women screened for FOB, 864 (19%) had a FOBS score of ≥60. Of 325 women who accepted participation in the RCT, 276 gave written informed consent. The main reasons for not willing to participate were that they felt no need for treatment (n=111), did not accept randomization (n=69), or felt that their fears (eg, fear of bleeding because of placenta previa, not finding an available hospital bed, or not reaching the hospital in time) could not be treated (n=61). In the end, 258 participants completed the preintervention questionnaire and were randomized as follows: 127 were allocated to the guided ICBT group and 131 to the SC group. Figure 1 shows the full Consolidated Standards of Reporting Trials flowchart.

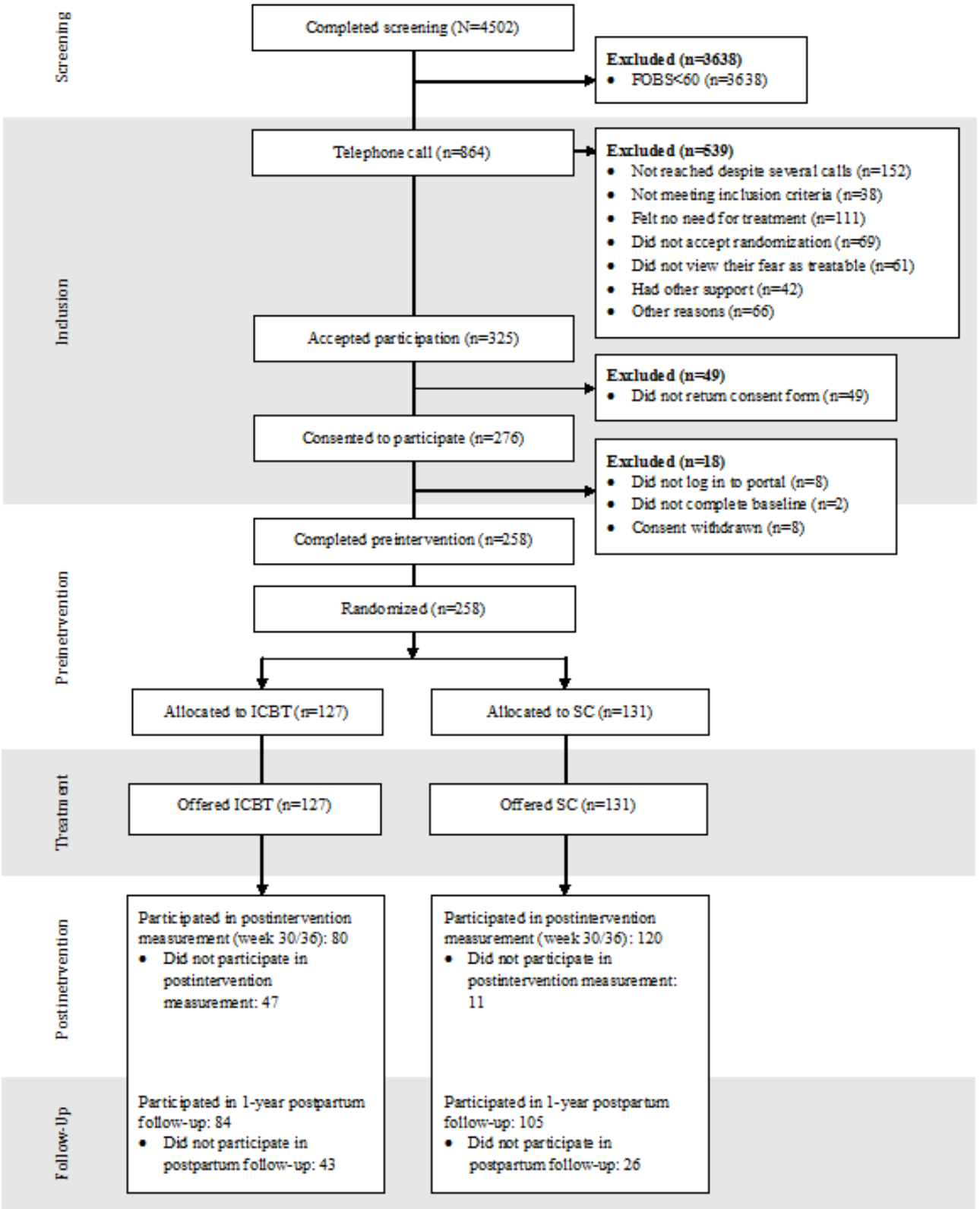
Guided Internet-Based Self-Help Program Based on Cognitive Behavioral Therapy

The aim of the guided ICBT intervention was to help participants observe and understand their FOB and find new ways of coping with difficult thoughts and emotions. With a study group that is likely to be highly heterogeneous (eg, with regard to parity and differences in fear acquisition, fear objects, symptom severity, and comorbidity), the treatment needed to be broad yet adequately flexible to be applicable to a wide range of different individual needs. Thus, the intervention was inspired by the unified protocol for transdiagnostic treatment of emotional disorders (UP), a broad face-to-face CBT protocol, designed for the applicability to all anxiety and unipolar mood disorders [46,47]. The study self-help material was built on the content of 7 of 8 modules in the UP, however, adapted to meet the needs of the current population (eg, with regard to the content and order of psychoeducative elements and by means of FOB-specific examples). Module 6 in the UP focuses specifically on the interoceptive exposure for induced symptoms. With a pregnant study sample, we decided to omit this module. To not risk getting into discussions of the accuracy of perceived threats, especially when not meeting participants face-to-face, we chose not to put too much emphasis on cognitive reappraisal (as presented in Module 4 in the UP). While still working on identifying automatic thoughts and giving some basic tools for reappraisal, we expanded the cognitive module by introducing exercises in cognitive defusion [48].

The self-help material was in Swedish and consisted of text material (81 downloadable PDF pages, including worksheets), audio files, photographs, and assignments related to each part of the program. The material was divided into 8 treatment modules and 1 module for the postpartum follow-up (see Textbox 1 for an overview). Participants were recommended to complete one self-help module per week. Each module

included 1-3 homework assignments that were reported using the internet-based platform. On completion of the assignments, participants received personalized written feedback and were given access to the next self-help module.

Figure 1. Flowchart of participants throughout the trial. ICBT: internet-based cognitive behavioral therapy SC: standard care.



Textbox 1. Overview of the guided internet-based self-help program based on cognitive behavioral therapy.

Introduction and motivation enhancement

- Introduction to the program
- Understanding fear and anxiety
- Motivation and behavioral change
- Assignment: Setting individual treatment goals

Emotion

- The function of emotion
- Physiological, cognitive, and behavioral aspects of emotion
- Assignment: Self-monitoring of emotional reactions

Behavior

- Learned and emotion-driven behaviors
- Avoidance and negative reinforcement
- Assignment: Self-monitoring of emotion-driven behaviors and avoidance behaviors

Cognition

- Automatic appraisals and catastrophizing
- Viewing cognitions as merely cognitions: working with cognitive defusion
- Assignment: Identification of childbirth-related catastrophic cognitions
- Assignment: Cognitive defusion exercises

Mindfulness and acceptance

- Nonjudgmental present-moment awareness
- Acceptance in relation to pregnancy and childbirth
- Assignment: Guided present-focused awareness
- Assignment: Anchoring in the present
- Assignment: Identifying childbirth- or pregnancy-related areas in need of acceptance

Exposure, part I

- The purpose and value of exposure-based interventions
- Different forms of exposure: situational, imaginative, and interoceptive
- Assignment: Generating a personalized avoidance hierarchy for emotional exposure

Exposure, part II

- Planning and implementation of exposure-based interventions
- Assignment: Exposure to images related to childbirth
- Assignment: Exposure to avoided situations in accordance with personal hierarchy

Generalization and maintenance

- Progress and acquired skills: a summary of the program
- Being your own therapist: working with maintenance, relapse prevention, and further development
- Assignment: Evaluation of personal progress and acquired skills
- Assignment: Creating a plan for maintenance and future development

Postpartum follow-up

- Childbirth in retrospect: the unique experience of each childbirth
- Generalizing acquired skills to other areas of life

- Assignment: Reviewing the childbirth experience: cognitions, emotions, and strategies
- Assignment: Exposure to images related to childbirth
- Assignment: How can the acquired skills be generalized to other areas of life?

The guided ICBT program was delivered through a secure internet-based platform, the U-CARE portal, using double verification for log-in. When randomized to the guided ICBT group, participants were also randomized to one of the two licensed clinical psychologists, who guided them through the self-help program. A welcome message was sent to each participant in the portal, along with a short message service (SMS) text message to their mobile phone. Participants who did not log in or follow the treatment plan received reminders, both in the portal and through SMS text messages, at 10 days and 4 weeks after randomization or their last log-in. About half-time through the project, the psychologists started to call each participant randomized to the guided ICBT group to optimize adherence and motivation. In total, 37 participants talked with their psychologist on the phone, whereas 15 did not respond despite several calls. The psychologists were active in the U-CARE portal three times a week, giving feedback on homework assignments, sending reminders, and answering messages from participants in the U-CARE portal.

Standard Care: Counseling by Midwives

All hospitals in Sweden provide SC for pregnant women with FOB [28]. Although guidelines exist [49], the content of SC and the time set aside for it differ between hospitals [28,50]. Women with FOB usually receive 2-4 counseling sessions either by antenatal midwives, counseling midwives and obstetricians, or a psychosocial unit consisting of midwives, obstetricians, and psychologists. The counseling aims at understanding the origin of fear, reducing the fear, preparing for childbirth, empowering women in their ability to give birth, and making the birth experience as positive as possible, regardless of the mode of birth [24]. Since SC is organized differently across the country, this also applies to SC at the study centers in this study. Depending on which study center a participant belonged to, SC started either in the next meeting with the antenatal midwife or after referral to a counseling midwife or a psychosocial unit [42].

Primary Outcome Measure

In this study, the primary outcome measure was the levels of FOB, measured in late pregnancy using FOBS [11,45]. This 2-item 100-mm visual analog scale consists of the question "How do you feel right now about the approaching birth," with the anchor words calm or worried and no fear or strong fear. FOBS has previously been used in child-bearing populations [45,51-53] and has been proposed as a valid instrument for measuring FOB in both research and clinical contexts [11,45]. The postpartum version of FOBS is worded as "How do you feel right now when thinking about giving birth again," with the anchor words calm or worried and no fear or strong fear.

Data Collection

Data were collected through self-assessment questionnaires at 4 time-points as follows: (1) at the ultrasound screening examination in gestational weeks 17-20; (2) through the U-CARE portal at preintervention in gestational weeks 20-25; (3) through the U-CARE portal at postintervention in gestational weeks 30 and 36; (4) through the U-CARE portal and offline questionnaires at follow-up, 1 year postpartum. Reminders were sent to each participant at 1, 6, 12, 30, and 38 days after the start of each time-point to maintain retention. Demographic and obstetric data were collected at preintervention. FOBS was included in all time-points.

Statistical Analysis

Statistical analyses were informed by the Consolidated Standards of Reporting Trials checklist [54] and conducted in the SPSS Statistics for Windows, version 24 (IBM Corp., Armonk, NY). Data from gestational weeks 30 and 36 were combined using the last observation carried forward, and the parity variable was dummy coded (0=primiparous, 1=multiparous). Participants who did not respond at either postintervention or follow-up were defined as lost to follow-up. Between-group differences in preintervention characteristics were analyzed using the independent-sample *t* test and Mann-Whitney *U* test for continuous variables and the Pearson's chi-square test for categorical variables. Little's missing completely at random test [55] was used to conclude that data were missing completely at random.

In this intention-to-treat study, linear mixed model analyses were used to analyze changes in FOB over time and whether such changes were dependent on the treatment allocation, parity of participants, or the interaction of both. Building on a likelihood-based approach, the linear mixed models analysis uses all available data and produces unbiased parameter estimates under the assumption of data being missing at random, making it suitable for intention-to-treat analyses in longitudinal studies with data missing at random [56-58]. We used the maximum likelihood estimation to compare the first basic model with subsequent models of increasing complexity using the likelihood ratio statistic [59].

The linear mixed model analyses were conducted in a sequence of nested models. The basic model examined the fixed effect of time on the dependent variable FOB, with a fixed intercept. The time variable represented the pregnancy week in which women responded to the questionnaires, with the intercept (point of zero) being the estimated due date. In the second model, a random effect of time, with a random intercept was included, adopting an unstructured covariance structure. Two different variants of the third model were conducted, one including the fixed effects of treatment and treatment × time, the other including the fixed effects of parity and parity × time. The fourth model involved all parameters included in the third models, as well as the three-way interaction between time, treatment, and parity. We compared the improved fit of each model with the preceding one using the likelihood ratio statistic [59].

As the linear mixed model analysis provides individual estimates of the outcome variable for each model tested, the estimated means were calculated from the values predicted in the last model that was statistically superior to prior models. We analyzed the between-group differences in FOB at postintervention and 1-year follow-up using the Mann-Whitney *U* test. Next, we calculated between- and within-group effect sizes (Cohen *d*) and their 95% CIs on the basis of both observed and estimated data. The clinically significant reduction in FOB was calculated and defined by a cutoff 2 SDs below the preintervention mean of the group [60]. Differences in the rate of treatment responders between the intervention groups were compared using the Pearson's chi-square test.

Results

Sample Characteristics

Table 1 presents the preintervention characteristics of study participants. The mean age of participants was 29.6 years (SD 4.88; range: 17-42 years), and 60% (154/258) of these were primiparous, whereas 40% (104/258) were multiparous. Regarding their preintervention characteristics, no difference was observed in the level of FOB between the parity groups. Primiparous women were younger (P<.001), and more often reported an eating disorder (P=.02), whereas multiparous women more often reported having had a previous miscarriage (P<.001) or abortion (P=.003; results not presented). Of the multiparous women, 36% (37/104) reported a previous negative birth experience, 22% (23/104) had experienced a previous emergency cesarean birth, and 25% (26/104) had experienced a birth aided by vacuum extraction. Of all participants, about 4% (10/258) were currently receiving CBT treatment, 11% (28/258) had participated in a CBT treatment prior to this pregnancy, and 7% (17/258) had received treatment for FOB prior to this pregnancy. The guided ICBT and SC groups did not differ with regard to any of the background characteristics or the level of FOB at screening or preintervention. Although all participants scored above the clinical cutoff for FOB at screening (FOBS ≥60), 52 (20%) scored below this cutoff at preintervention.

Treatment Adherence

Table 2 shows the number of treatment modules opened by the participants in the guided ICBT group. Of all participants allocated to this intervention, 81% (103/127) commenced treatment. Among these, the mean time logged in the portal was 39.96 minutes (SD 49.88; range: 1-244 minutes) or 13.21 minutes per opened module (SD 10.03; range: 0.5-47). Primiparous and multiparous women did not differ with regard to any of the variables related to the treatment adherence. Feedback regarding the adherence to SC could not be retrieved from care providers. All participants randomized to the SC group did not report whether they received SC. Of 79 women responding to this question, 3 (4%) reported not having participating in any treatment. In accordance with the intention-to-treat principle, all participants were asked to complete postintervention and follow-up assessments, regardless of the treatment adherence.

Missing Data Analysis

The Little's missing completely at random test showed that data were missing completely at random in the primary outcome variable (χ^2_8=9.8, P=.28). Further analysis showed that participants defined as lost to follow-up (did not respond either at postintervention or follow-up) were no different from the other participants with regard to any preintervention characteristic or the level of FOB at screening or preintervention. However, participants lost to follow-up were more likely to belong to the guided ICBT group (χ^2_1=11.2, P<.001). Overall, 24 (18.9%) participants in the ICBT group and 7 (5.3%) in the SC group were lost to follow-up.

Descriptive Statistics, Mean Differences, and Effect Sizes

Figure 2 plots and Tables 3 and 4 present the observed and estimated means and SDs of the primary outcome measure, along with the within-group effect sizes (Cohen *d*) and 95% CIs. The estimated means were calculated from the individual values of FOB predicted in Model 3a in the linear mixed model analysis. The levels of FOB did not differ between the intervention groups at postintervention. At 1-year postpartum follow-up, participants in the guided ICBT group exhibited significantly lower levels of FOB, both in the observed and estimated data (U=3674.00, z=–1.97, P=.049 and U=6985.00, z=–2.23, P=.027, respectively). Although the within-group effect sizes were generally found to be moderate or large, the between-group effect sizes were small or very small, Cohen d=0.14 favoring SC at postintervention and Cohen d=0.28 favoring the guided ICBT at follow-up in the observed data. The estimated between-group effect sizes were Cohen d=0.15 at postintervention, and Cohen d=0.29 at follow-up. At postintervention, 99 of 200 responding women had a FOBS score of ≥60 (guided ICBT group, 44/80; SC group, 55/120). At follow-up, the corresponding figures were 65 of 189 in the total sample (guided ICBT group, 29/84; SC group, 36/105).

Responder Analysis

In line with recommendations by Jacobson and Truax [60], the cutoff for responding to treatment was set at 2 SDs below the preintervention mean of the group (FOBS≤38). At postintervention, a significantly higher proportion of participants in the SC group scored below this cutoff, 29 (22.1%) compared with 11 (8.7%) in the guided ICBT group (χ^2_1=8.9, P=.003). At follow-up, the groups did not differ significantly, with 44 (34.6%) participants in the guided ICBT group and 37 (28.2%) participants in the SC group reaching below this cutoff.

Linear Mixed Model Analysis

The basic linear mixed model showed a significant effect of time on FOB. Overall, the FOBS score decreased from screening to follow-up ($F_{1,905}$=220.08, P<.001).

Table 1. Characteristics of participants at preintervention.

Characteristics	Guided internet-based cognitive behavioral therapy (n=127), n (valid %)	Standard care (n=131), n (valid %)	All participants (n=258), n (valid %)
Age in years			
<25	21 (16.5)	16 (12.2)	37 (14.3)
25-35	90 (70.9)	96 (73.3)	186 (72.1)
>35	16 (12.6)	19 (14.5)	35 (13.6)
Civil status			
Living with partner	121 (95.3)	122 (93.1)	243 (94.2)
Not living with partner	6 (4.7)	9 (6.9)	15 (5.8)
Level of education			
Compulsory school or high school	55 (43.3)	65 (49.6)	120 (46.5)
University education	72 (56.7)	66 (50.4)	138 (53.5)
Country of birth			
Sweden	108 (85.0)	116 (88.5)	224 (86.8)
Other country	19 (15.0)	15 (11.5)	34 (13.2)
Computer illiterate			
Yes	5 (3.9)	4 (3.1)	9 (3.5)
No	122 (96.1)	127 (96.9)	249 (96.5)
Previous abortion			
Yes	31 (24.4)	30 (22.9)	61 (23.6)
No	96 (75.6)	101 (77.1)	197 (76.4)
Previous miscarriage			
Yes	30 (23.6)	30 (22.9)	60 (23.3)
No	97 (76.4)	101 (77.1)	198 (76.7)
Ongoing or history of depression			
Yes	41 (32.3)	50 (38.2)	91 (35.3)
No	86 (67.7)	81 (61.8)	167 (64.7)
Ongoing or history of anxiety			
Yes	35 (27.6)	40 (30.5)	75 (29.1)
No	92 (72.4)	91 (69.5)	183 (70.9)
Ongoing or history of an eating disorder			
Yes	14 (11.0)	16 (12.2)	30 (11.6)
No	113 (89.0)	115 (87.8)	228 (88.4)
Ongoing or history of bipolar disorder			
Yes	0 (0.0)	3 (2.3)	3 (1.2)
No	127 (100.0)	128 (97.7)	255 (98.8)
Ongoing or history of other psychiatric disorder			
Yes	7 (5.5)	15 (11.5)	22 (8.5)
No	120 (94.5)	116 (88.5)	236 (91.5)
Using medication for depression or anxiety at present			
Yes	7 (5.5)	13 (9.9)	20 (7.8)
No	120 (94.5)	118 (90.1)	238 (92.2)

Table 2. Participants in the guided internet-based cognitive behavioral therapy (ICBT) group who opened each treatment module and the mean time spent per module.

Module	ICBT group (n=127)		Primiparas (n=77)		Multiparas (n=50)	
	Opened module n (%)	Minutes in module[a] Mean	Opened module n (%)	Minutes in module[a] Mean	Opened module n (%)	Minutes in module[a] Mean
1	103 (81)	12.49	64 (83)	12.56	39 (78)	12.36
2	60 (47)	10.97	36 (47)	11.17	24 (48)	10.67
3	35 (28)	29.49	23 (30)	31.00	12 (24)	26.58
4	24 (19)	22.38	19 (25)	22.47	5 (10)	22.00
5	13 (10)	35.62	10 (13)	34.80	3 (6)	38.33
6	7 (6)	12.86	5 (6)	9.60	2 (4)	21.00
7	1 (1)	29.00	0 (0)	0.00	1 (2)	29.00
8	1 (1)	20.00	0 (0)	0.00	1 (2)	20.00
9[b]	1 (1)	1.00	0 (0)	0.00	1 (2)	1.00

[a]Mean time, measured in minutes, spent per module by participants who opened the module.

[b]Module for the postpartum follow-up.

Table 3. Observed means and SDs of Fear of Birth Scale scores at screening, preintervention, postintervention, and follow-up, including the within-group effect sizes.

Type of Intervention	Descriptive statistics		Effect size	
	n	Mean (SD)	Cohen *d*	95% CI
Guided ICBT[a]				
Screening	127	74.76 (10.38)		
Preintervention	127	74.06 (16.70)		
Postintervention	80	60.56 (21.63)	0.58[b]	0.26-0.89
Follow-up	84	41.17 (32.65)	1.23[c]	0.89-1.55
Standard Care				
Screening	131	74.96 (11.36)		
Preintervention	131	71.44 (17.99)		
Postintervention	120	57.20 (24.83)	0.70[d]	0.44-0.96
Follow-up	105	50.11 (30.48)	0.86[e]	0.58-1.14

[a]ICBT: internet-based cognitive behavioral therapy.

[b]Preintervention (n=80) mean 71.80 (SD 16.88).

[c]Preintervention (n=84) mean 72.92 (SD 16.49).

[d]Preintervention (n=120) mean 72.20 (SD 17.47).

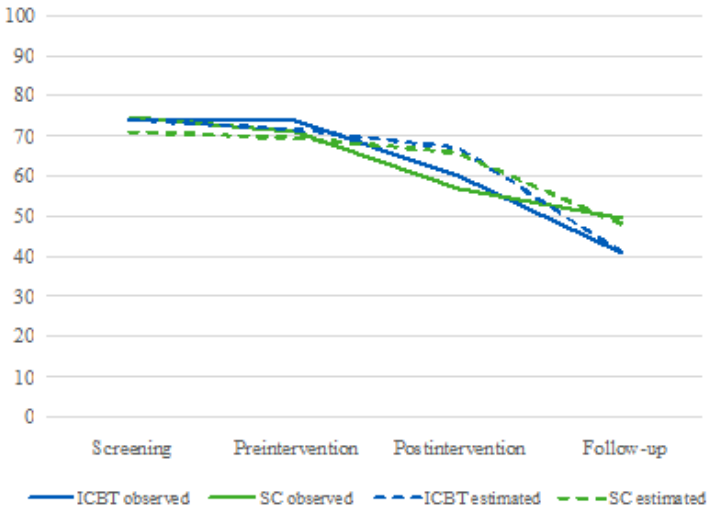
[e]Preintervention (n=80) mean 71.58 (SD 17.69).

Table 4. Estimated means and SDs of the Fear of Birth Scale scores at screening, preintervention, postintervention, and follow-up, including the within-group effect sizes.

Type of Intervention	Descriptive statistics		Effect size	
	n	Mean (SD)	Cohen *d*	95% CI
Guided ICBT[a]				
Screening	127	74.26 (5.73)		
Preintervention	127	71.76 (6.61)		
Postintervention	127	67.15 (8.62)	0.60	0.35-0.85
Follow-up	127	41.03 (22.45)	1.86	1.56-2.14
Standard Care				
Screening	131	70.94 (6.33)		
Preintervention	131	69.29 (7.33)		
Postintervention	131	65.73 (9.65)	0.42	0.17-0.66
Follow-up	131	47.87 (24.10)	1.20	0.94-1.46

[a]ICBT: internet-based cognitive behavioral therapy.

Figure 2. Observed and estimated Fear of Birth Scale mean scores from screening until 1 year postpartum. ICBT: internet-based cognitive behavioral therapy; SC: standard care.



The second model, examining whether the effect of time on FOB differed between individuals, showed significantly better fit with data than the first model (χ^2_3=214.4, *P*<.001). Significant variance remaining in the intercept and time variable indicated significant differences between participants with regard to the estimated level of FOB at the estimated due date and in the effect of time on FOB. Overall, this implies that both levels of FOB and how these levels changed over time differed significantly between participants.

In the third pair of models, we investigated whether these individual differences in the effect of time on FOB could be attributed to the treatment allocation or parity. In Model 3a, a significant interaction between treatment and time was found ($F_{1,192.538}$=4.96, *P*=.03), showing that the reduction in FOB over time was significantly larger in the guided ICBT group (−0.46 units/week) than in the SC group (−0.31 units/week). However, the predicted level of FOB at the estimated due date did not differ significantly (guided ICBT group, 64.61; SC group, 64.15; $t_{1,240.996}$=−0.24, *P*=.81). Hence, when comparing the intervention groups, no difference was observed in FOB in late pregnancy. When considering the entire study period, FOB decreased more in the guided ICBT group. In comparison with Model 2, Model 3a showed significantly better fit with data (χ^2_2=7.8, *P*=.02). However, Model 3b did not show a better fit with data than Model 2 (χ^2_2=0.5, *P*=.80), with no significant interaction between parity and time. Hence, changes in FOB over time were not significantly different between primiparous and multiparous women.

The main purpose of Model 4 was to examine the possibility of a three-way interaction between time, treatment, and parity. No such interaction effect was found, and Model 4 did not show a better fit with data compared with Model 3a (χ^2_4=0.5, *P*=.97), suggesting that the interaction between time and treatment did not differ depending on parity.

Discussion

Principal Findings

In this study, the level of FOB was found to decrease over time in both groups, generally with medium within-group effect sizes during pregnancy and large effect sizes from midpregnancy to 1 year postpartum. Similar decreases in FOB over time have been shown before, both among women with mixed levels of fear in early pregnancy [11] and among women receiving support for FOB [27]. This apparent effect of time alone points to the importance of including a proper control group when evaluating treatments for FOB. It is, thus, possible that the previously shown within-group effect of internet-based, therapist-supported, self-help for FOB [41] could, at least, in part, be attributable to a natural decrease in FOB over time. Unexpectedly, in 20% of participants, we observed a reduction of FOB below the inclusion criteria cutoff already before randomization and the introduction of any planned intervention. Although possibly an effect of the passage of time alone, this reduction might also be related to participants talking to a research midwife on the phone to be included in the study or simply because of statistical regression to the mean. As pregnancy itself is time-limited and the utmost feared situation will ultimately occur, the passage of time might have a unique meaning within this particular population.

When comparing the different interventions, participants allocated to the SC group were more likely to have responded to the treatment at postintervention measurement. However, mean differences were not significant at this time, and the between-group effect size was ignorable (Cohen *d*=0.14). In the linear mixed model analyses, a small, yet significant, interaction between time and treatment was found, indicating that over time FOB decreased slightly more among participants allocated to the guided ICBT group. This effect was most evident at 1-year postpartum follow-up, when participants in the guided ICBT group exhibited significantly lower levels of FOB, however, still with a small between-group effect size (Cohen *d*=0.28). This finding is not easily interpreted. First, given the low adherence to treatment in the guided ICBT group, these differences might not be attributable to the ICBT intervention. Perhaps, the differences rather relate to the interventions provided in the SC group. If so, our results might be in line with recently published results showing that women who receive SC for FOB still have higher levels of FOB in late pregnancy than women with FOB who do not receive SC [11]. Although highly valued by women receiving this form of support [5,61], the design of interventions provided in SC might be more focused on reducing FOB during the ongoing pregnancy (eg, by means of being able to convey what feels important during birth or planning for pain relief) than on treating fear in a long-term perspective.

Second, as the difference between the intervention groups does not appear until 1 year postpartum, multiple factors during childbirth and in the postpartum period might mediate this effect. Although findings are not coherent [62], previous research suggests that FOB during pregnancy might be positively correlated with the experience of pain during childbirth [63,64], longer birth duration [65-67], and a more negative rating of the overall birth experience [4,5]. Some studies have suggested a higher number of emergency cesarean births among women with FOB [68-70]. Furthermore, postpartum levels of FOB have been associated with previous mode of delivery, intervention at birth, or emergency cesarean births [4,13,20], as with more negative birth experiences [4,13]. Overall, when trying to understand FOB in the postpartum period, outcomes and experiences of giving birth are likely to contribute significantly, either as mediators, moderators, or as confounding variables. In this study, randomization and the resulting equivalence between the intervention groups prior to intervention will contribute to the minimization of the effect of extraneous variables.

In this trial, both primiparous and multiparous women were included. Although these groups are commonly separated and assumed to be in need of different interventions, there is nothing in our results that points specifically to that conclusion. We did not find any difference in FOB or the effect of the different interventions between primiparas and multiparas. Moreover, no difference was observed with regard to the treatment adherence or participant dropout. Hence, as far as we can see, none of these treatment alternatives seems to suit either parity group better.

Limitations

This study has several methodological limitations—the most problematic of these related to the poor treatment adherence, participants being lost to follow-up, and wide inclusion criteria giving room for sample heterogeneity.

Concerning the SC group, we have no information on who actually received any counseling, how many appointments each participant had, who conducted the counseling or what it consisted of. We can only rely on the results of Larsson et al [28], showing that counseling exists nationwide but differs considerably in aspects such as available treatment options for women and educational background and time set aside for health care professionals providing counseling.

In the guided ICBT group, very low treatment adherence is obvious. Unfortunately, it is difficult to know all the reasons participants had for not engaging in their ICBT. Some participants reported reduced levels of fear, changes in their life circumstances, having received other forms of treatment, not having sufficient time, or problems related to the internet-based portal, whereas most did not respond to any attempt from the study team to get in contact. In this study, we did not measure the treatment acceptance or credibility. However, it seems likely to assume that the guided ICBT was not a well-accepted intervention in this sample. Quite a few potential participants declined participation because they did not accept randomization to either intervention, that is, they preferred SC beforehand. The 24 participants who did not commence the guided ICBT at all are likely to belong to that group as well. Fewer than half of the participants allocated to the guided ICBT group finished the first module and went on to the second module, and less than one-third advanced to Module 3. This poor adherence can be attributed to several possible reasons, potentially related to expectations and care preferences in the population, the process

of inclusion and exclusion of participants, lack of preintervention assessment of individual needs, issues related to the U-CARE portal, instructions and reminders not being sufficient, treatment format, or self-help material not meeting the expectations of participants.

Besides low treatment adherence, there were quite large amounts of missing data, resulting in the need to combine two postintervention measures. Unfortunately, the amount of missing data was particularly evident in the guided ICBT group, presumably because of the low treatment adherence and participants having difficulties in differentiating between their treatment and data collection. However, with data being missing completely at random, we could use all available data and perform the intention-to-treat analysis using linear mixed models with the maximum likelihood estimation.

Contribution

Although this study has some apparent limitations, it also has strengths. First, to the best of our knowledge, this is the first study using a randomized controlled design to evaluate the effects of CBT on FOB. The randomized controlled design and the equivalence between the intervention groups are important factors enhancing the internal validity of a study [71]. The inclusion of a control group is important to differentiate between the effect of the intervention and confounding variables. Although wait-list controls are commonly used in psychotherapy research, the limited time of pregnancy makes this control condition difficult to apply with regard to FOB. Although SC is difficult to control and thus may threaten the internal validity of the study, we still found this the most appropriate control condition available. Importantly, it gave us the possibility to differentiate the effect of treatment with what appears to be the effect of time alone.

Second, although primarily an efficacy study, the generous inclusion of participants and the naturalistic setting of this study resemble the prioritizations of effectiveness studies [71], resulting in a reduced level of control regarding potential confounding variables (eg, heterogeneity within the sample in terms of the symptom severity, comorbidity, and concurrent treatments). Despite these problems, the results still give a hint of how guided ICBT could work in a naturalistic setting. Since participants were included when visiting standard antenatal care, they have not actively asked for treatment for their FOB. Hence, the sample is not likely to be comparable with highly motivated samples of participants who have self-recruited to guided ICBT, as in the study by Nieminen et al [41] and many studies evaluating CBT delivered over the internet [72]. Instead, with regard to their age, civil status, and level of education, participants in this study were very similar to the general birthing population in Sweden [73]. As women with insufficient knowledge in the Swedish language were excluded from this study, the results cannot be generalized to this population.

Conclusions

In this study, FOB decreased during pregnancy and until 1 year postpartum, both in the guided ICBT group and the SC group. The reduction in FOB was similar in the intervention groups during pregnancy, and the effect of time alone appeared as more important than the specific effect of any intervention. One year postpartum, a stronger reduction in FOB was found in the guided ICBT group—a finding that was not easily interpreted given the low adherence to the guided ICBT and the wide array of potential mediators, moderators, and confounders during childbirth and the postpartum period. Hence, the guided ICBT, as offered in this study, did not seem to be a feasible or well-accepted approach for treating FOB.

Future Directions

The challenge in future research will be to find an intervention that is both well accepted by pregnant women and effective in reducing FOB. Considering the strong evidence for CBT in treating anxiety, cognitive and behavioral interventions should not be ruled out at this early point. However, to enhance the credibility among pregnant women and caregivers, we need to learn more about the experiences of women participating in different intervention programs and make adjustments in the treatment format and structure based on their views. Instead of comparing different treatment interventions, it might be more fruitful to integrate existing and well-accepted midwife counseling with CBT interventions, acknowledging the need for individual tailoring.

Acknowledgments

The authors would sincerely like to thank the participating women, study centers, recruiting midwives, research assistants, and U-CARE research team for their participation and support. This study was awarded funding by Uppsala University and Uppsala University Hospital and is part of a published thesis.

Authors' Contributions

ER and ET conducted the statistical analyses and drafted the manuscript. CR, IH, HMH, JE, and ÖS provided substantial revision. All authors have reviewed the text and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 826KB - mental_v5i3e10420_app1.pdf]

References

1. Heimstad R, Dahloe R, Laache I, Skogvoll E, Schei B. Fear of childbirth and history of abuse: implications for pregnancy and delivery. Acta Obstet Gynecol Scand 2006;85(4):435-440. [doi: 10.1080/00016340500432507] [Medline: 16612705]
2. Laursen M, Hedegaard M, Johansen C. Fear of childbirth: predictors and temporal changes among nulliparous women in the Danish National Birth Cohort. BJOG 2008 Feb;115(3):354-360 [FREE Full text] [doi: 10.1111/j.1471-0528.2007.01583.x] [Medline: 18190372]
3. Molgora S, Fenaroli V, Prino LE, Rollè L, Sechi C, Trovato A, et al. Fear of childbirth in primiparous Italian pregnant women: The role of anxiety, depression, and couple adjustment. Women Birth 2017 Jul 11;31(2):117-123. [doi: 10.1016/j.wombi.2017.06.022] [Medline: 28709777]
4. Handelzalts J, Becker G, Ahren M, Lurie S, Raz N, Tamir Z, et al. Personality, fear of childbirth and birth outcomes in nulliparous women. Arch Gynecol Obstet 2015 May;291(5):1055-1062. [doi: 10.1007/s00404-014-3532-x] [Medline: 25391637]
5. Larsson B, Karlström A, Rubertsson C, Hildingsson I. The effects of counseling on fear of childbirth. Acta Obstet Gynecol Scand 2015 Jun;94(6):629-636. [doi: 10.1111/aogs.12634] [Medline: 25772528]
6. Karlström A, Rådestad I, Eriksson C, Rubertsson C, Nystedt A, Hildingsson I. Cesarean section without medical reason, 1997 to 2006: a Swedish register study. Birth 2010 Mar;37(1):11-20. [doi: 10.1111/j.1523-536X.2009.00373.x] [Medline: 20402717]
7. Sydsjö G, Sydsjö A, Gunnervik C, Bladh M, Josefsson A. Obstetric outcome for women who received individualized treatment for fear of childbirth during pregnancy. Acta Obstet Gynecol Scand 2012 Jan;91(1):44-49. [doi: 10.1111/j.1600-0412.2011.01242.x] [Medline: 21787365]
8. O'Connell MA, Leahy-Warren P, Khashan AS, Kenny LC, O'Neill SM. Worldwide prevalence of tocophobia in pregnant women: systematic review and meta-analysis. Acta Obstet Gynecol Scand 2017 Aug;96(8):907-920. [doi: 10.1111/aogs.13138] [Medline: 28369672]
9. Rondung E, Thomtén J, Sundin Ö. Psychological perspectives on fear of childbirth. J Anxiety Disord 2016 Dec;44:80-91 [FREE Full text] [doi: 10.1016/j.janxdis.2016.10.007] [Medline: 27788373]
10. Huizink AC. Prenatal stress and its effect on infant development Internet. Utrecht: Utrecht University Repository; 2000. URL: https://dspace.library.uu.nl/bitstream/handle/1874/371/inhoud.htm [accessed 2018-07-19] [WebCite Cache ID 712E5sjzK]
11. Hildingsson I, Haines H, Karlström A, Nystedt A. Presence and process of fear of birth during pregnancy-Findings from a longitudinal cohort study. Women Birth 2017 Oct;30(5):e242-e247. [doi: 10.1016/j.wombi.2017.02.003] [Medline: 28279636]
12. Lukasse M, Vangen S, Øian P, Schei B. Fear of childbirth, women's preference for cesarean section and childhood abuse: a longitudinal study. Acta Obstet Gynecol Scand 2011 Jan;90(1):33-40. [doi: 10.1111/j.1600-0412.2010.01024.x] [Medline: 21275913]
13. Nilsson C, Lundgren I, Karlström A, Hildingsson I. Self reported fear of childbirth and its association with women's birth experience and mode of delivery: a longitudinal population-based study. Women and Birth 2012 Sep;25(3):114-121. [doi: 10.1016/j.wombi.2011.06.001] [Medline: 21764400]
14. Nilsson C, Hessman E, Sjöblom H, Dencker A, Jangsten E, Mollberg M, et al. Definitions, measurements and prevalence of fear of childbirth: a systematic review. BMC Pregnancy Childbirth 2018 Dec 12;18(1):28 [FREE Full text] [doi: 10.1186/s12884-018-1659-7] [Medline: 29329526]
15. Saisto T, Halmesmaki E. Fear of childbirth: a neglected dilemma. Acta Obstet Gynecol Scand 2003 Mar;82(3):201-208. [doi: 10.1034/j.1600-0412.2003.00114.x]
16. Areskog B, Uddenberg N, Kjessler B. Fear of childbirth in late pregnancy. Gynecol Obstet Invest 1981;12(5):262-266. [doi: 10.1159/000299611] [Medline: 7262639]
17. Wijma K, Wijma B, Zar M. Psychometric aspects of the W-DEQ; a new questionnaire for the measurement of fear of childbirth. Journal of Psychosomatic Obstetrics & Gynecology 2009 Jul 07;19(2):84-97. [doi: 10.3109/01674829809048501]
18. Hofberg K, Brockington I. Tokophobia: An unreasoning dread of childbirth. Br J Psychiatry 2000;176(01):83-85. [doi: 10.1192/bjp.176.1.83]
19. Huizink A, Mulder E, Robles de Medina P, Visser G, Buitelaar J. Is pregnancy anxiety a distinctive syndrome? Early Hum Dev 2004 Sep;79(2):81-91. [doi: 10.1016/j.earlhumdev.2004.04.014] [Medline: 15324989]
20. Fenwick J, Gamble J, Nathan E, Bayes S, Hauck Y. Pre- and postpartum levels of childbirth fear and the relationship to birth outcomes in a cohort of Australian women. J Clin Nurs 2009 Mar;18(5):667-677. [doi: 10.1111/j.1365-2702.2008.02568.x] [Medline: 19239535]
21. Melender HL. Experiences of Fears Associated with Pregnancy and Childbirth: A Study of 329 Pregnant Women. Birth 2002 Jun;29(2):101-111. [doi: 10.1046/j.1523-536X.2002.00170.x]

22. Rouhe H, Salmela-Aro K, Gissler M, Halmesmäki E, Saisto T. Mental health problems common in women with fear of childbirth. BJOG 2011 Aug;118(9):1104-1111 [FREE Full text] [doi: 10.1111/j.1471-0528.2011.02967.x] [Medline: 21489127]
23. Sjögren B. Reasons for anxiety about childbirth in 100 pregnant women. Journal of Psychosomatic Obstetrics & Gynecology 2009 Jul 07;18(4):266-272. [doi: 10.3109/01674829709080698]
24. Rouhe H, Salmela-Aro K, Toivanen R, Tokola M, Halmesmäki E, Saisto T. Obstetric outcome after intervention for severe fear of childbirth in nulliparous women - randomised trial. BJOG 2013 Jan;120(1):75-84 [FREE Full text] [doi: 10.1111/1471-0528.12011] [Medline: 23121002]
25. Rouhe H, Salmela-Aro K, Toivanen R, Tokola M, Halmesmäki E, Ryding E, et al. Group psychoeducation with relaxation for severe fear of childbirth improves maternal adjustment and childbirth experience--a randomised controlled trial. J Psychosom Obstet Gynaecol 2015 Nov;36(1):1-9. [doi: 10.3109/0167482X.2014.980722] [Medline: 25417935]
26. Saisto T, Toivanen R, Salmela-Aro K, Halmesmäki E. Therapeutic group psychoeducation and relaxation in treating fear of childbirth. Acta Obstet Gynecol Scand 2006 Jan;85(11):1315-1319. [doi: 10.1080/00016340600756920] [Medline: 17091410]
27. Toohill J, Fenwick J, Gamble J, Creedy D, Buist A, Turkstra E, et al. A randomized controlled trial of a psycho-education intervention by midwives in reducing childbirth fear in pregnant women. Birth 2014 Dec;41(4):384-394 [FREE Full text] [doi: 10.1111/birt.12136] [Medline: 25303111]
28. Larsson B, Karlström A, Rubertsson C, Hildingsson I. Counseling for childbirth fear - a national survey. Sex Reprod Healthc 2016 Jun;8:82-87. [doi: 10.1016/j.srhc.2016.02.008] [Medline: 27179382]
29. National Institute for Heath and Care Excellence. Social anxiety disorder: recognition, assessment and treatment. NICE clinical guideline 159. 2013. URL: https://www.nice.org.uk/guidance/cg159 [accessed 2018-07-20] [WebCite Cache ID 713F0Du4W]
30. National Institute for Heath and Care Excellence. Anxiety disorders. NICE quality standard 53. 2014. URL: https://www.nice.org.uk/guidance/qs53 [accessed 2018-07-20] [WebCite Cache ID 713FNIJ0v]
31. National Institute for Health and Clinical Excellence. Obsessive-compulsive disorder and body dysmorphic disorder: treatment. NICE clinical guideline 31. 2005. URL: https://www.nice.org.uk/guidance/cg31 [accessed 2018-07-20] [WebCite Cache ID 713Fcb27u]
32. National Institute for Health and Clinical Excellence. Post-traumatic stress disorder: management. NICE clinical guideline 26. 2005. URL: https://www.nice.org.uk/guidance/cg26 [accessed 2018-07-20] [WebCite Cache ID 713FjwV1Y]
33. National Institute for Health and Clinical Excellence. Generalized anxiety disorder and panic disorder in adults: management. NICE clinical guideline 113. 2011. URL: https://www.nice.org.uk/guidance/cg113 [accessed 2018-07-20] [WebCite Cache ID 713FoxRtO]
34. National Institute for Health and Clinical Excellence. Depression in adults: recognition and management. NICE clinical guideline 90. 2009. URL: https://www.nice.org.uk/guidance/cg90 [accessed 2018-07-20] [WebCite Cache ID 713G7ovpd]
35. National Institute for Health and Care Excellence. Antenatal and postnatal mental health: clinical management and service guidance. NICE clinival guideline 192. 2014. URL: https://www.nice.org.uk/guidance/cg192 [accessed 2018-07-20] [WebCite Cache ID 713GCIMtL]
36. Pearl S, Norton P. Transdiagnostic versus diagnosis specific cognitive behavioural therapies for anxiety: A meta-analysis. J Anxiety Disord 2017 Mar;46:11-24. [doi: 10.1016/j.janxdis.2016.07.004] [Medline: 27466074]
37. 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 May;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
38. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
39. Andrews G, Cuijpers P, Craske M, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
40. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
41. Nieminen K, Andersson G, Wijma B, Ryding E, Wijma K. Treatment of nulliparous women with severe fear of childbirth via the Internet: a feasibility study. J Psychosom Obstet Gynaecol 2016;37(2):37-43. [doi: 10.3109/0167482X.2016.1140143] [Medline: 26918752]
42. Ternström E, Hildingsson I, Haines H, Karlström A, Sundin Ö, Ekdahl J, et al. A randomized controlled study comparing internet-based cognitive behavioral therapy and counselling by standard care for fear of birth - A study protocol. Sex Reprod Healthc 2017 Oct;13:75-82. [doi: 10.1016/j.srhc.2017.06.001] [Medline: 28844361]
43. U-CARE. URL: https://www.u-care.se/ [accessed 2018-03-02] [WebCite Cache ID 6xcDC4mLG]
44. Hildingsson I, Nilsson C, Karlström A, Lundgren I. A longitudinal survey of childbirth-related fear and associated factors. J Obstet Gynecol Neonatal Nurs 2011;40(5):532-543. [doi: 10.1111/j.1552-6909.2011.01274.x] [Medline: 22273410]

45. Haines H, Pallant J, Karlström A, Hildingsson I. Cross-cultural comparison of levels of childbirth-related fear in an Australian and Swedish sample. Midwifery 2011 Aug;27(4):560-567. [doi: 10.1016/j.midw.2010.05.004] [Medline: 20598787]
46. Barlow DH. Unified protocol for transdiagnostic treatment of emotional disorders: therapist guide. New York: Oxford University Press; 2011.
47. Ellard K, Fairholme C, Boisseau C, Farchione T, Barlow D. Unified Protocol for the Transdiagnostic Treatment of Emotional Disorders: Protocol Development and Initial Outcome Data. Cognitive and Behavioral Practice 2010 Feb;17(1):88-101. [doi: 10.1016/j.cbpra.2009.06.002]
48. Harris R. ACT made simple: an easy-to-read primer on acceptance and commitment therapy. Oakland, CA: New Harbinger Publications; 2009.
49. Björklund U, Ellis J, Baummann Gustafsson B, Hammarström M, Lilliecreutz C, Nieminen K, et al. Svensk Förening för Obstetrik och Gynekologi [Swedish Society of Obstetrics and Gynecology]. 2017. Förlossningsrädsla [Fear of Childbirth] URL: https://www.sfog.se/natupplaga/1005623_A2fb46497-d45b-466d-8126-c58e99fe856d.pdf [accessed 2018-07-20] [WebCite Cache ID 713JiKBdm]
50. Ryding EL, Persson A, Onell C, Kvist L. An evaluation of midwives' counseling of pregnant women in fear of childbirth. Acta Obstet Gynecol Scand 2003 Jan;82(1):10-17. [doi: 10.1034/j.1600-0412.2003.820102.x]
51. Haines HM, Pallant JF, Fenwick J, Gamble J, Creedy DK, Toohill J, et al. Identifying women who are afraid of giving birth: A comparison of the fear of birth scale with the WDEQ-A in a large Australian cohort. Sex Reprod Healthc 2015 Dec;6(4):204-210. [doi: 10.1016/j.srhc.2015.05.002] [Medline: 26614602]
52. Ternström E, Hildingsson I, Haines H, Rubertsson C. Higher prevalence of childbirth related fear in foreign born pregnant women--findings from a community sample in Sweden. Midwifery 2015 Apr;31(4):445-450. [doi: 10.1016/j.midw.2014.11.011] [Medline: 25529841]
53. Ternström E, Hildingsson I, Haines H, Rubertsson C. Pregnant women's thoughts when assessing fear of birth on the Fear of Birth Scale. Women Birth 2016 Jun;29(3):e44-e49. [doi: 10.1016/j.wombi.2015.11.009] [Medline: 26710973]
54. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. Ann Intern Med 2017 Jul 04;167(1):40-47. [doi: 10.7326/M17-0046] [Medline: 28630973]
55. Little R. A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association 1988 Dec;83(404):1198-1202. [doi: 10.2307/2290157]
56. Gueorguieva R, Krystal J. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
57. Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm Stat 2008 Apr;7(2):93-106. [doi: 10.1002/pst.267] [Medline: 17351897]
58. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data (Springer Series in Statistics). New York: Springer; 2000.
59. Gumedze F, Dunne T. Parameter estimation and inference in the linear mixed model. Linear Algebra and its Applications 2011 Oct;435(8):1920-1944. [doi: 10.1016/j.laa.2011.04.015]
60. Jacobson N, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 1991;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12]]
61. Larsson B, Karlström A, Rubertsson C, Ternström E, Ekdahl J, Segebladh B, et al. Birth preference in women undergoing treatment for childbirth fear: A randomised controlled trial. Women Birth 2017 Dec;30(6):460-467. [doi: 10.1016/j.wombi.2017.04.004] [Medline: 28495462]
62. Garthus-Niegel S, Knoph C, von Soest T, Nielsen C, Eberhard-Gran M. The role of labor pain and overall birth experience in the development of posttraumatic stress symptoms: a longitudinal cohort study. Birth 2014 Mar;41(1):108-115. [doi: 10.1111/birt.12093] [Medline: 24654643]
63. Beebe K, Lee K, Carrieri-Kohlman V, Humphreys J. The effects of childbirth self-efficacy and anxiety during pregnancy on prehospitalization labor. J Obstet Gynecol Neonatal Nurs 2007;36(5):410-418. [doi: 10.1111/j.1552-6909.2007.00170.x] [Medline: 17880311]
64. Kjærgaard H, Wijma K, Dykes A, Alehagen S. Fear of childbirth in obstetrically low‐risk nulliparous women in Sweden and Denmark. Journal of Reproductive and Infant Psychology 2008 Nov;26(4):340-350. [doi: 10.1080/02646830802408498]
65. Adams S, Eberhard-Gran M, Eskild A. Fear of childbirth and duration of labour: a study of 2206 women with intended vaginal delivery. BJOG 2012 Sep;119(10):1238-1246 [FREE Full text] [doi: 10.1111/j.1471-0528.2012.03433.x] [Medline: 22734617]
66. Reck C, Zimmer K, Dubber S, Zipser B, Schlehe B, Gawlik S. The influence of general anxiety and childbirth-specific anxiety on birth outcome. Arch Womens Ment Health 2013 Oct;16(5):363-369. [doi: 10.1007/s00737-013-0344-0] [Medline: 23558948]
67. Sydsjö G, Angerbjörn L, Palmquist S, Bladh M, Sydsjö A, Josefsson A. Secondary fear of childbirth prolongs the time to subsequent delivery. Acta Obstet Gynecol Scand 2013 Feb;92(2):210-214. [doi: 10.1111/aogs.12034] [Medline: 23066797]

68. Hildingsson I. Swedish couples' attitudes towards birth, childbirth fear and birth preferences and relation to mode of birth - a longitudinal cohort study. Sex Reprod Healthc 2014 Jun;5(2):75-80. [doi: 10.1016/j.srhc.2014.02.002] [Medline: 24814442]
69. Laursen M, Johansen C, Hedegaard M. Fear of childbirth and risk for birth complications in nulliparous women in the Danish National Birth Cohort. BJOG 2009 Sep;116(10):1350-1355 [FREE Full text] [doi: 10.1111/j.1471-0528.2009.02250.x] [Medline: 19538412]
70. Ryding E, Lukasse M, Van Parys AS, Wangel A, Karro H, Kristjansdottir H, Bidens Group. Fear of childbirth and risk of cesarean delivery: a cohort study in six European countries. Birth 2015 Mar;42(1):48-55. [doi: 10.1111/birt.12147] [Medline: 25676793]
71. Nathan PE, Stuart SP, Dolan SL. Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? Psychological Bulletin 2000;126(6):964-981. [doi: 10.1037/0033-2909.126.6.964]
72. Andersson G, Bergström J, Buhrman M, Carlbring P, Holländare F, Kaldo V, et al. Development of a New Approach to Guided Self-Help via the Internet: The Swedish Experience. Journal of Technology in Human Services 2008 Jul 03;26(2-4):161-181. [doi: 10.1080/15228830802094627]
73. Socialstyrelsen [The National Board of Health and welfare]. Graviditeter, Förlossningar och Nyfödda Barn, Medicinska Födelseregistret 1973–2014, Assisterad Befruktning 1991-2013 [Pregnancies, Deliveries and Newborn Infants, The Swedish Medical Birth Register 1973-2014, Assisted Reproduction Treatment 1991-2013]. 2015. URL: http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/20009/2015-12-27.pdf [accessed 2018-05-08] [WebCite Cache ID 713Lwf6GS]

Abbreviations

CBT: cognitive behavioral therapy
FOB: fear of birth
FOBS: Fear of Birth Scale
ICBT: internet-based cognitive behavioral therapy
RCT: Randomized controlled trial
SC: standard care
U-CARE: Uppsala University Psychosocial Care Program
UP: unified protocol

Edited by J Torous; submitted 19.03.18; peer-reviewed by S Garthus-Niegel, M Berg; comments to author 29.04.18; revised version received 21.05.18; accepted 06.06.18; published 10.08.18.

Please cite as:
Rondung E, Ternström E, Hildingsson I, Haines HM, Sundin Ö, Ekdahl J, Karlström A, Larsson B, Segeblad B, Baylis R, Rubertsson C
Comparing Internet-Based Cognitive Behavioral Therapy With Standard Care for Women With Fear of Birth: Randomized Controlled Trial
JMIR Ment Health 2018;5(3):e10420
URL: http://mental.jmir.org/2018/3/e10420/
doi:10.2196/10420
PMID:30097422

Original Paper

Predicting Social Anxiety From Global Positioning System Traces of College Students: Feasibility Study

Mehdi Boukhechba[1], PhD; Philip Chow[2], PhD; Karl Fua[2], PhD; Bethany A Teachman[2], PhD; Laura E Barnes[1], PhD

[1]Systems and Information Engineering Department, School of Engineering and Applied Science, University of Virginia, Charlottesville, VA, United States

[2]Department of Psychology, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Laura E Barnes, PhD
School of Engineering and Applied Science
Systems and Information Engineering Department
University of Virginia
151 Engineer's Way, Olsson Hall 101B
Charlottesville, VA, 22904
United States
Phone: 1 (434) 924 1723
Email: lb3dp@virginia.edu

Abstract

Background: Social anxiety is highly prevalent among college students. Current methodologies for detecting symptoms are based on client self-report in traditional clinical settings. Self-report is subject to recall bias, while visiting a clinic requires a high level of motivation. Assessment methods that use passively collected data hold promise for detecting social anxiety symptoms and supplementing self-report measures. Continuously collected location data may provide a fine-grained and ecologically valid way to assess social anxiety in situ.

Objective: The objective of our study was to examine the feasibility of leveraging noninvasive mobile sensing technology to passively assess college students' social anxiety levels. Specifically, we explored the different relationships between mobility and social anxiety to build a predictive model that assessed social anxiety from passively generated Global Positioning System (GPS) data.

Methods: We recruited 228 undergraduate participants from a Southeast American university. Social anxiety symptoms were assessed using self-report instruments at a baseline laboratory session. An app installed on participants' personal mobile phones passively sensed data from the GPS sensor for 2 weeks. The proposed framework supports longitudinal, dynamic tracking of college students to evaluate the relationship between their social anxiety and movement patterns in the college campus environment. We first extracted the following mobility features: (1) cumulative staying time at each different location, (2) the distribution of visits over time, (3) the entropy of locations, and (4) the frequency of transitions between locations. Next, we studied the correlation between these features and participants' social anxiety scores to enhance the understanding of how students' social anxiety levels are associated with their mobility. Finally, we used a neural network-based prediction method to predict social anxiety symptoms from the extracted daily mobility features.

Results: Several mobility features correlated with social anxiety levels. Location entropy was negatively associated with social anxiety (during weekdays, $r=-0.67$; and during weekends, $r=-0.51$). More (vs less) socially anxious students were found to avoid public areas and engage in less leisure activities during evenings and weekends, choosing instead to spend more time at home after school (4 pm-12 am). Our prediction method based on extracted mobility features from GPS trajectories successfully classified participants as high or low socially anxious with an accuracy of 85% and predicted their social anxiety score (on a scale of 0-80) with a root-mean-square error of 7.06.

Conclusions: Results indicate that extracting and analyzing mobility features may help to reveal how social anxiety symptoms manifest in the daily lives of college students. Given the ubiquity of mobile phones in our society, understanding how to leverage passively sensed data has strong potential to address the growing needs for mental health monitoring and treatment.

(JMIR Ment Health 2018;5(3):e10101) doi:10.2196/10101

KEYWORDS

mental health; mHealth; mobility; GPS; social anxiety disorder

Introduction

Social anxiety is marked by an extreme fear of being scrutinized and judged by others in social or performance situations [1]. In addition to being a widespread problem among college students, a high social anxiety level is associated with a low quality of life. For example, socially anxious individuals suffer from impaired academic functioning and relationships [2]. The American College Health Association reported that 40% of students reported feeling "overwhelming anxiety" at least once in the preceding year [3].

Current techniques to identify social anxiety are typically based on self-report via questionnaires or interviews in traditional clinical settings, where only small numbers of people can be monitored and client motivation is required to seek an assessment. This approach is inadequate and fails to meet the growing needs of mental health monitoring and treatment on college campuses. As a result, many individuals who should seek help never receive any. For example, according to the Anxiety and Depression Association of America, 36% people with social anxiety disorder report having symptoms for 10 or more years before seeking help [1].

Mental health symptoms can be indirectly assessed through both subjective (eg, self-report surveys and interviews) and objective (physiological variables such as heart rate) methods. Such methods have largely been employed in clinical or laboratory settings, which limits the ecological validity of findings. To increase the generalizability of findings, researchers have tried to understand mental health through noninvasive and real-time data collected from people's everyday lives. For example, studies using surveys to repeatedly sample people's momentary affective experiences over time have found that individuals with high (vs low) social anxiety symptoms report more anger and fewer and less intense positive emotions [4,5]. While studies that administer repeated surveys offer a glimpse of the socioemotional aspects of daily life, regularly prompting individuals to answer questions also raises the issue of participation burden.

Importantly, embedded mobile phone sensors (eg, accelerometers, light sensors, Global Positioning System [GPS]) are now advanced enough to allow for passive and continuous data collection [6,7] and are increasingly being used to enhance the understanding of the relationship between objective behavior and mental health status, such as bipolar disorder [8], anxiety and depression [7,9-13], and Alzheimer disease and dementia [14]. Digital phenotyping, which is a term used for describing this new approach of measuring behavior from mobile phones and wearable sensors, is already revealing new aspects of behaviors that appear clinically relevant [15]. For example, Saeb et al [16,17] provided preliminary evidence that extracting location-based mobility features could be used to detect depression and anxiety levels. Barnet et al [18] were able to use passively generated mobile phone data to identify statistically significant anomalies in patients with schizophrenic behavior in the days prior to relapse. The above studies show the importance of analyzing behaviors to better understand mental states.

Because social anxiety is marked by intense fear of social scrutiny, passively sensed location information may reveal key markers that can be used to detect a high distress level. Semantic locations (ie, the type of social location someone visits) might be particularly important in the context of social anxiety. For example, individuals with social anxiety might systematically avoid specific places, such as those of leisure, or choose to spend peak social hours by isolating themselves at home. Thus, analyzing GPS data from college students and the types of places they frequent might provide crucial information about key mobility features associated with social anxiety levels. Some examples of mobility features include how long students spend at different types of locations (eg, home, leisure) and how often they frequent those locations.

Important contributions [7,9,12,17,19] have been made to determine how passively sensed mobile phone data relates to users' mental health statuses and stress levels. We followed the key steps from these valuable early studies and extended the scope of features and questions addressed, as outlined below: (1) recruiting participants and deploying a mobile app for data collection; (2) collecting data such as GPS locations, recognized activities, or self-reported data from participants during the study; (3) assessing participants' mental health status using validated clinical measures; (4) extracting meaningful features or metrics from participants' data (eg, time spent at each different location); and (5) correlating these features or metrics with participants' mental health status (eg, Pearson correlation between number of distinct locations and clinically validated measures).

This study builds on prior work in several ways and improves our understanding of the relationships between social anxiety symptoms and daily routines. In this paper, we introduce and analyze a set of passively extracted spatiotemporal features that enhance our understanding of the temporal and spatial dynamics of behavioral patterns (eg, regularly visiting a location during specific hours) of socially anxious students, which may allow for more precise, personalized interventions. We also propose a hierarchical social anxiety prediction method based on neural networks. This work may ultimately help researchers and clinicians to passively and remotely monitor patients' social anxiety levels. The primary aim of this paper was not to test specific hypotheses, but rather to explore a framework for using passively collected GPS data to detect social anxiety levels.

Methods

Study Organization and Data Collection

Participants were undergraduate students with varying social anxiety levels, recruited from undergraduate psychology classes that provide course credit as a participation incentive. Because some participants met or exceeded their course credit limit, a

subset of participants was eligible to receive a small amount of monetary compensation (up to a maximum of US $40). Students were recruited through email advertisements as well as through an undergraduate study participant pool. The decision to recruit university students was based on two reasons: (1) there are high social anxiety levels among young adults, and (2) recruiting young adults in a university setting provides a relatively homogeneous sample in terms of life phase, psychological stressors, and life experiences, thereby eliminating a wide variety of potential confounding factors.

After receiving approval from the university Institutional Review Board, 228 participants were recruited. Social anxiety level was first assessed via the Social Interaction Anxiety Scale (SIAS) [20] in a prestudy screening battery offered to select undergraduate psychology classes in exchange for course credit. The SIAS contains 20 items, each rated from 0 to 4. Generally, a higher SIAS score (specifically, higher than 34) [21] indicates a higher risk of having social anxiety concerns; a low score indicates a lower risk for social anxiety concerns.

Following informed consent, a custom mobile app (Sensus) [22] was installed on participants' personal mobile phone (either IOS or Android device). As shown in Figure 1, participants were informed that the app passively collected the GPS location information every 150 seconds and uploaded it to an Amazon Web Services S3 server. After the 2-week experiment was completed, researchers could access all participants' raw GPS data for analysis.

Global Positioning System Data Preprocessing

Participants' raw GPS data were parsed by semantic locations (eg, restaurant, campus area, and shops) by combining a spatiotemporal clustering algorithm and OpenStreetMap (OSM) geodatabase [23]. Specifically, we first clustered participants' GPS locations using time and space dimensions, and then, each cluster was associated with a semantic location using OSM data [24] (Figure 2).

Our clustering algorithm is inspired by the work of Kang et al [25], and it aims to eliminate the intermediate locations between important places and determine the number of clusters (important places) autonomously. The core idea guiding our approach is to cluster the locations along the time axis. As a new location measurement is read, the new location is compared with previous locations. If the new location is moving away from previous locations (within a certain distance of each other—a parameter of our algorithm), the new location is considered to belong to a different cluster than the one for the previous locations. Otherwise, it is considered to belong to the previous cluster. If a cluster's time duration is longer than the threshold (the second parameter of our algorithm), the cluster is considered to be a significant place (see A and B in Figure 2); otherwise, it is ignored (see i1 and i2 in Figure 2).

The algorithm is depicted in Textbox 1 (*d* and *t* are our distance and time threshold parameters). When a new location measurement event is generated, the cluster function is invoked. The current cluster *cl* is the set of location measurements that belong to the current cluster. The pending location *pl* is used to eliminate outliers. Even if the new location is far away from the current cluster (distance is larger than the distance threshold d), the algorithm does not start a new cluster right away with the new location. Instead, the algorithm waits for the next location to determine if the user is really moving away from the cluster or if the location reading was just a spurious outlier. The *Places* contain the significant places where the user stays longer than the time threshold *t*.

Figure 1. Social anxiety monitoring framework.

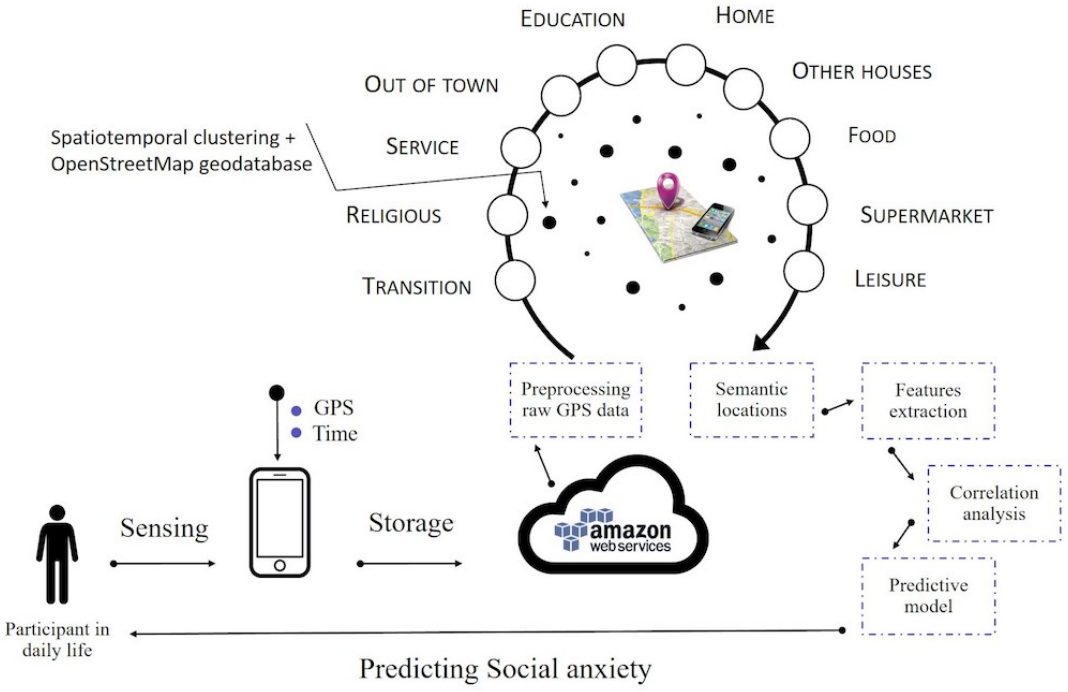
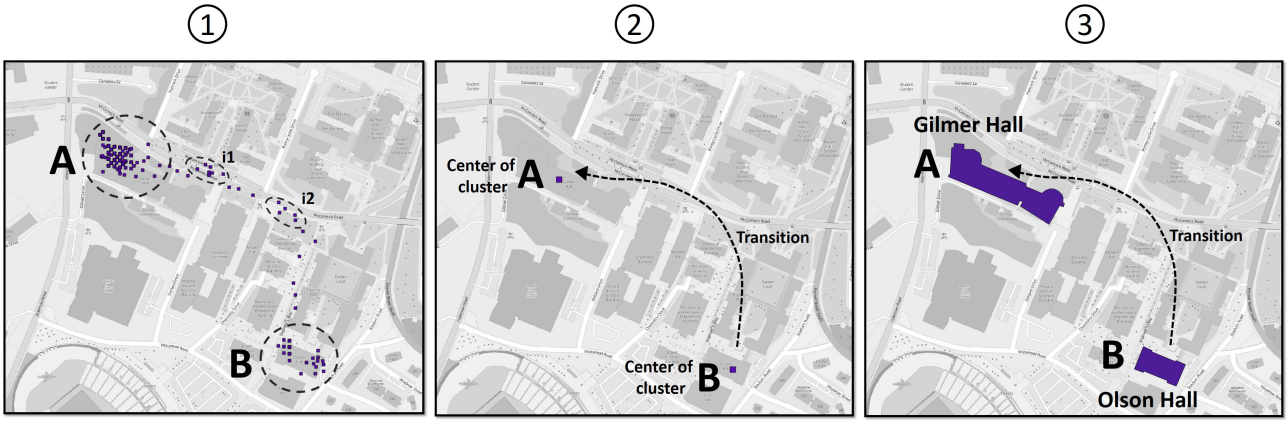


Figure 2. Illustration of our time-based clustering algorithm using real GPS data retrieved from one participant in the study. In (1), GPS locations are clustered using the algorithm described in Textbox 1. In (2), the trajectory is summarized to two places (A and B) and the transition state, which aggregates all GPS points between the clusters A and B. Finally, in (3), the clusters A and B are labeled using OSM data. GPS: Global Positioning System, OSM: open street map.



Textbox 1. Spatiotemporal clustering algorithm.

Input: measured location *loc*

Output: current cluster *cl*, pending location *pl*, significant places *Places*

if distance (*cl*, *loc*) < d **then**

add *loc* to *cl*

pl = null

else

if*pl* ≠ null **then**

if duration(*cl*) > *t* then

if contain long gaps(*cl*) **then**

remove gaps from *cl*

end if

add *cl* to *Places*

end if

clear *cl*

add *pl* to *cl*

if distance (*cl*, *pl*) < *d***then**

add *loc* to *cl*

pl = null

else

pl = *loc*

end if

else

pl = *loc*

end if

end if

Our algorithm was tuned using d=60 m and t=600 seconds. These values appeared to give the best clustering results for our data; they have also been reported in the literature to give the best performance for spatiotemporal clustering algorithms [25].

After detecting the significant clusters, we transformed each GPS cluster to a meaningful semantic label using OSM data. Each GPS cluster's centroid is associated to a geographic entity (eg, in Figure 2, cluster A is associated to Gilmer Hall and cluster B to Olsson Hall, both of which are buildings on the

university campus) using a spatial query in our geodatabase powered by OSM. The semantic data obtained from OSM is then classified to one of the following classes:

- **Home**: our algorithm has been trained to recognize "Home" as the place having a house OSM tag (eg, apartment, dormitory, house, etc) where a subject stays the most between 10 pm and 9 am; see [23] for more details about OSM tags.
- **Other houses**: all houses other than "Home"; in this study, given all participants are college students, other houses were assumed to mostly be friends' houses.
- **Education**: eg, university and libraries
- **Leisure**: eg, sports locations, pubs, cinemas, and coffee shops
- **Food**: eg, dining halls and restaurants including fast food joints
- **Supermarket**: all full-service grocery stores
- **Religious**: all places of worship, including churches, mosques, cathedrals, synagogues, temples, etc
- **Service**: eg, bank, post office, courthouse
- **Out of town**: locations outside of the city where the study was conducted
- **In transition**: going from one place to another

Note that ideally, food places would be merged with leisure and supermarkets with service classes. However, we decided to separate them because we discovered a particular pattern that high socially anxious participants (SAP) share in terms of time spent at food places and supermarkets, which will be discussed in the next section.

When constructing GPS clusters labeled with semantics, we verified if the users' data contained GPS gaps. We defined a GPS gap g_i as a minimum 10-minute time span where GPS data were missing. Gaps could be caused by different events, such as turning the phone off or "killing" the app. For gaps $\in$ [5 min, 30 min], we compared the cluster cl_i and the cluster cl_{i+1}, created before and after the gap, respectively. If the 2 clusters had the same semantic labels, we considered that the user did not change his or her location during g_i; thus, we merged the clusters cl_i, cl_{i+1}, and the gap g_i. However, if the 2 clusters had different semantic labels, we assigned the "Transition" label to g_i; ie, the user changed locations during this gap.

For gaps exceeding 30 minutes in duration, we removed the corresponding time periods from the experiment (see Textbox 1, line 7), because it was hard to predict what the participants did during such long gaps.

Mobility Feature Extraction

After detecting participants' visited places and labeling them using OSM, we used the semantic labels to identify the following mobility features for each participant:

Cumulative Staying Time in each type of location: Given a type of location and a specific participant, this metric characterized the percentage of total time that the participant spent at one type of location during a specific time window (eg, during a day, during mornings vs afternoons).

Distribution of visits for each type of location: Given a type of location and a specific participant, this metric calculated the density distribution of time of visits over the time of day. For instance, if a participant was more likely to go to leisure places during evenings, we would find more density during the evening periods for this type of locations. We introduced this metric because cumulative staying time captures only time spent at each different location without recording when these visits happened; for instance, spending 2 hours at the university during mornings was different than that during evenings.

Location entropy: This metric was calculated using the entropy of Shannon [26] to measure how each participant's time was distributed over different location classes, where p_i is the percentage of time spent at location i and n is the total number of visited locations:

Transition Frequency from one type of location to another: Given a specific participant and two types of locations (eg, "Home" and "Work"), this metric characterized the frequency at which the participant transited from one type of location to another. This metric was applied unidirectionally; for example, the transition frequency of "Home" "Work" and the transition frequency of "Work" "Home" were different.

Results

Participants

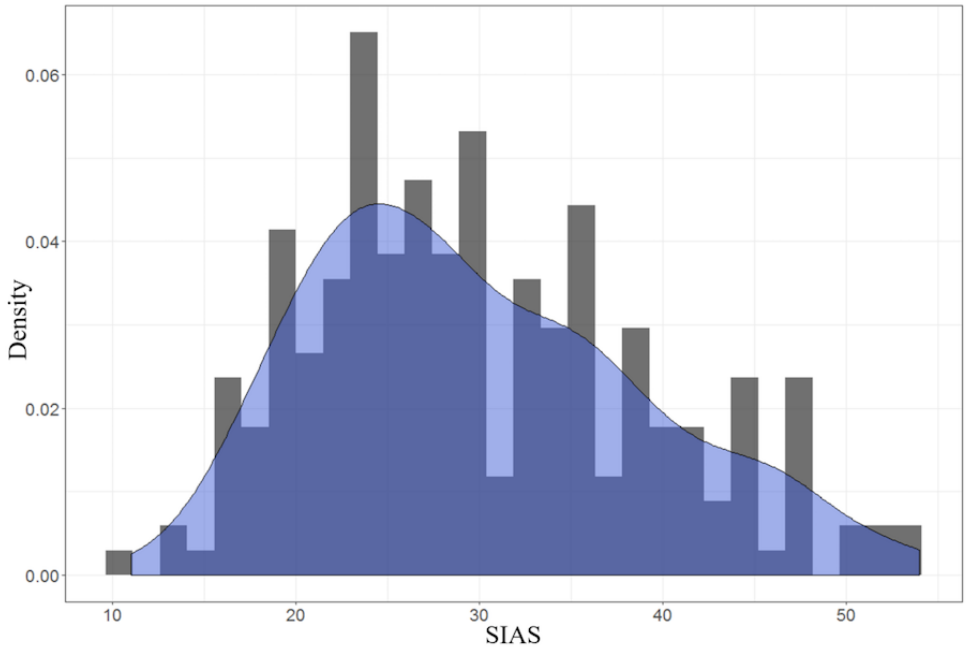
Participants comprised 228 university students (mean age 19.43 [SD 2.92] years; 141/228, 62%, females). Participants reported their race or ethnicity as white (118/228, 52%), Asian (61/228, 27%), black (12/228, 5%), Latino (5/228, 2%), and multiracial (32/228, 14%). Figure 3 shows the distribution of SIAS scores among the 228 participants. The SIAS scores ranged between 11 and 54 with a mean score 29.91 (SD 9.1).

The goal of our study was: (1) to understand the relationship between the extracted mobility features and the preassessed SIAS score and (2) to investigate whether the extracted features could predict SIAS scores. Consequently, in the next two sections, we will first analyze the relationship between the mobility features and social anxiety and then investigate whether these features can accurately predict social anxiety.

Mobility Data Analysis

In this section, we present the results of our analysis investigating the relationship between social anxiety levels (using the preassessed SIAS measures) and the four extracted mobility metrics: cumulative staying time, distribution of visits, location entropy, and transition frequency.

Figure 3. The distribution of Social Interaction Anxiety Scale (SIAS) scores for recruited participants. The Epanechnikov kernel function was used to compute the density estimates presented in this figure.



Cumulative Staying Time

We calculated the Pearson correlation between each participant's average daily cumulative staying time at each different location and his or her SIAS score to identify the direction (positive or negative) and strength of each correlation. To assess the reliability of the correlations, we also calculated significance levels (*P* value).

Results presented in graph 1 of Figure 4 show that daily time spent at some locations was associated with the SIAS score. Indeed, time spent at food locations, such as restaurants and dining halls, was negatively correlated with the SIAS score. However, time spent at supermarkets was positively correlated with the SIAS score. This suggests that high SAP are more likely to buy food from supermarkets so they can eat at home, perhaps to avoid social interactions at restaurants.

College students may have common mobility patterns that bias the daily correlation analysis, such as class times at the university following a typical schedule. In order to find the hidden patterns between the cumulative staying time and SIAS score, we analyzed correlations in different daily time epochs: 8 am-4 pm, 4 pm-12 am, and 12 am-8 am. Results presented in graphs 2, 3, and 4 of Figure 4 suggest the following:

- Similar to the 24-hour analysis (Figure 4, graph 1), the time spent at food locations was negatively correlated with the SIAS score, while time spent at supermarkets was positively correlated with the SIAS score in both the 12 am-8 am and 8 am-4 pm time windows.
- Time spent doing leisure activities was positively correlated with the SIAS score between 8 am and 4 pm, while the rest of time it was negatively correlated with the SIAS score. This suggests that high SAP prefer to do their leisure activities between 8 am and 4 pm, rather than during the evenings. This may reflect the social demands typical of different types of leisure activities done during the day versus evening (eg, it is more normative to be alone at a coffee shop than at a pub or bar).
- We did not find a correlation between time spent at home and the SIAS score between 12 am and 4 pm (Figure 4, graphs 2 and 4), which may simply indicate that no matter how socially anxious students were, they tended to stay at home (sleeping) between 12 am-8 am and leave home to go to school between 8 am and 4pm. However, during the time after typical school hours (between 4 pm and 12 am, when students have the choice to stay at home or not), we found a positive correlation between the SIAS score and time spent at home (Figure 4, graph 3). This finding is consistent with a prior work that associates time spent at home with depressive and social isolation symptoms [27].
- Finally, we found a small correlation (0.22) between time spent out of town and the SIAS score during the 4 pm-12 am time window, perhaps indicating that more socially anxious students leave the university to visit familiar individuals or family, rather than engaging in campus night life, which may have more demands to be social with unfamiliar individuals.

After analyzing the correlation between cumulative staying time and the SIAS score, we studied the difference between cumulative staying time during weekdays versus weekends for high (SIAS score ≥ 34) versus low (SIAS < 34) SAP. A score of 34 is an established clinical cutoff for the SIAS score to classify a subject as high or low socially anxious [21]. The reason for this analysis is that students' patterns may differ between weekdays and weekends. For instance, maybe the time spent at the university is not a good predictor of social anxiety during weekdays, but it is during weekends when students presumably have more autonomy in determining their schedule (because classes are not set).

Figure 4. The correlations between time spent at each different type of location and Social Interaction Anxiety Scale (SIAS) scores using different time windows. In (1), we have presented the correlations on a daily basis, while in the other figures, we have focused on specific portions of the day, ie, between 8 am and 4 pm, between 4 pm and 12 am, and between 12 am and 8 am. The x-axis represents the correlation significance; the left y-axis describes the type of locations, and the right y-axis represents the P value of the Pearson correlation in that specific type of location. Correlations having a coefficient $r>0.2$ and a strength $P<0.05$ are presented in purple.

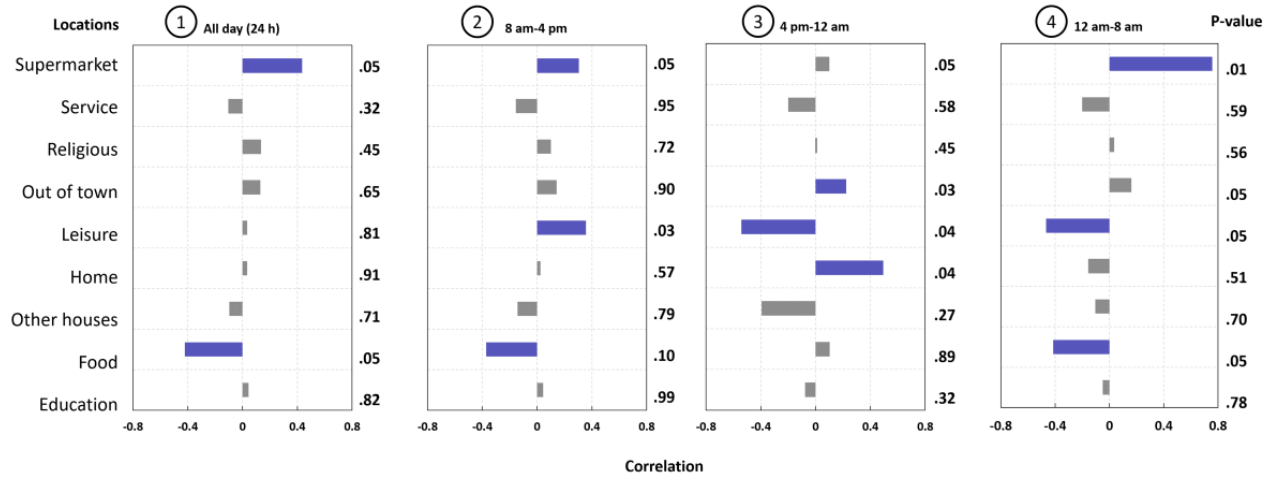


Table 1. The difference between high and low socially anxious students in terms of average daily time spent (in minutes) at each different location during weekdays versus weekends.

Location	Weekdays		Weekends	
	High[a]	Low[a]	High[a]	Low[a]
Supermarket	6.64[b]	1.41[b]	16.64[b]	10.41[b]
Service	6.97	1.83	9.97	5.73
Religious	7.35	1.63	10.35	2.75
Out of town	38.25[b]	22.33[b]	88.25[b]	27.8[b]
Leisure	26.04[b]	45.19[b]	43.16[b]	75.09[b]
Home	434.74	386.94	594.74[b]	469.02[b]
Other houses	17.63[b]	24.81[b]	16.21[b]	64.81[b]
Food	3.78[b]	12.97[b]	9.78[b]	32.97[b]
Education	338.44	316.88	148.79[b]	102.98[b]

[a]Social anxiety levels were classified to high and low using SIAS score=34 as cutoff.

[b]Significant differences ($P<.05$) between high and low SAP detected using unpaired two-samples *t* test.

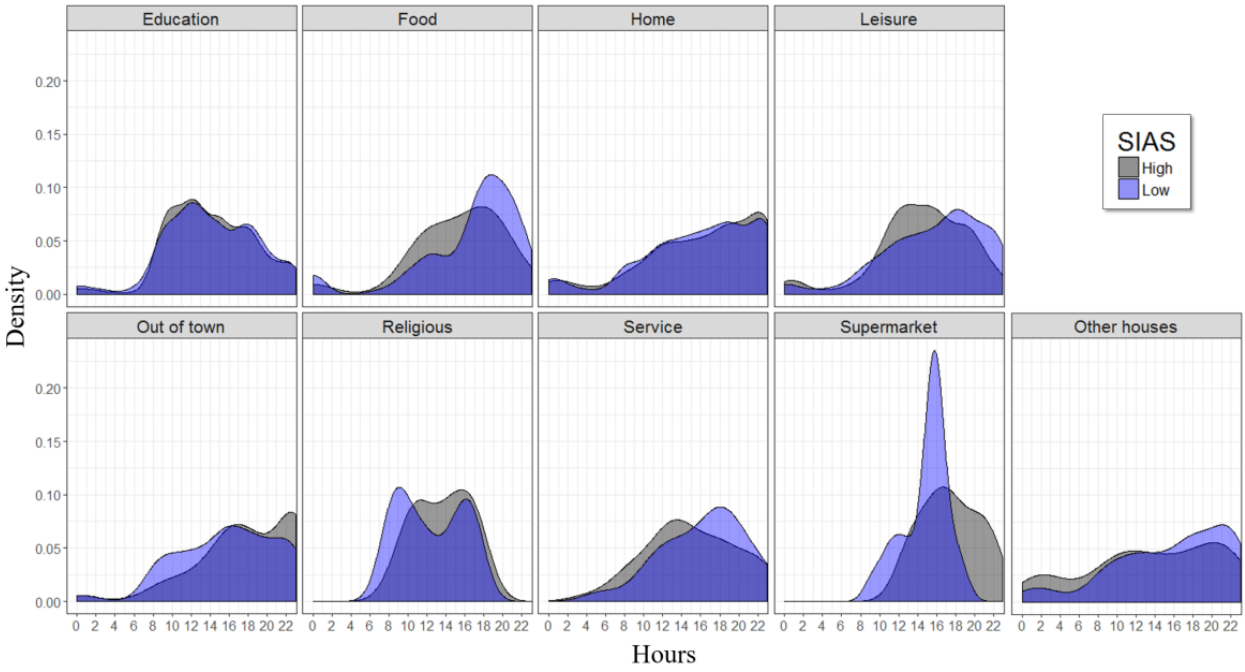
Results presented in Table 1 show that high SAP spent less time at leisure and food places and more time at home and out of town during both weekdays and weekends. However, during weekends, high SAP tended to spend more time at education places (around 148 minute) compared with low SAP (around 102 minute). They also appeared to spend less time at other houses (a difference of 48 minute), perhaps because they were less comfortable engaging in social interactions that may happen at friends' houses or simply had fewer opportunities (invitations) for these interactions.

Distribution of Visits

To better understand the relationship between students' daily routines and their social anxiety, we analyzed the distribution of location visits over the day for both high and low SAP. Figure 5 illustrates the difference in the distribution of location visits between the two populations. Note that this figure analyzes the time of visits (the time that a participant visited a specific location) without considering the duration of visits because cumulative staying time has already been studied above.

The results show a difference in the pattern of visits to food places, supermarkets, others' houses, and leisure places. Low SAP appeared to prefer going to friends' houses and food and leisure places during evenings (after 4 pm) more than high SAP. On the other hand, high SAP preferred to stay at home or go to the supermarket during that time period. This suggests that there may be a difference in how students plan their daily activities based on how socially anxious they are. Understanding these patterns may help clinicians identify when a person is starting to withdraw more socially and both plan and easily monitor specific social activity targets in treatment.

Figure 5. The distribution of location visits over the day for both high (grey) and low (blue) socially anxious participants. The Epanechnikov kernel function was used to compute the density estimates presented in this figure. SIAS: Social Interaction Anxiety Scale.



For instance, it is common during cognitive behavioral therapy for social anxiety disorder to have clients plan novel social activities, both to have more opportunities to receive reinforcement from the environment and reduce withdrawal and to engage in previously avoided activities to learn that the social anxiety can be tolerated (these "exposures" are designed to reduce social avoidance and provide new learning opportunities). Having clients identify good opportunities for these social activities and monitoring what they did over the past week can be challenging; thus, this real-time monitoring could help with treatment planning and assessing progress in ecologically valid ways in real time.

Location Entropy

We calculated the correlation between location entropy and the SIAS score during weekdays and weekends. Results showed that the diversity of places visited on both weekdays and weekends was negatively correlated with the SIAS score (r=−0.64, P=.001 for weekdays and r=−0.57, P=.001 for weekends), which suggests that socially anxious students visit fewer *different* places and have a narrower range of activities. This finding is consistent with a previous work [7] concluding that location entropy is strongly related to feelings of sadness and stress among students. Importantly, tracking the range of places a person with social anxiety visits could be a very useful marker of treatment progress because increasing the range may indicate an expanded repertoire of social skills and engagement, given different activities require different types of social interactions (eg, attending a party at a house is quite different from having dinner on a date with one person). It also likely indicates a willingness to try new social activities.

Frequency of Transitions

Next, we investigated whether transition frequency between each different location correlated with SIAS scores. Table 2 shows the results of the Pearson correlation (with P values) between transition frequencies and SIAS scores. The total number of distinct transitions was 152. In Table 2, we have presented only transitions having a correlation coefficient higher than 0.3 and P value <.05. Results show that some transitions correlated positively with SIAS scores, such as going from education to supermarket, while other transitions correlated negatively with SIAS scores, such as going from one leisure place to another leisure place. These data provide convergent support for the earlier findings, suggesting that socially anxious individuals are less likely to do leisure activities later in the day (thus, do not transition from one leisure activity to another) and more likely to do more typically solitary activities like grocery shopping.

Studying transition periods may be useful clinically to the extent they represent potential instances of approach-avoidance conflict in socially anxious individuals [4]. A hallmark feature of social anxiety is avoidance of settings in which social scrutiny is likely to occur [28]. Therefore, individuals may be less likely to transition from socially secure locations to insecure locations and more likely to transition from socially insecure locations to secure locations. In addition, the frequency of transitions might be a meaningful metric, such that more transitions are made on days with a higher anxiety level. Studying transitions could, thus, ultimately allow for greater insight into decision making about approach and avoidance behaviors in the real world.

Predicting Social Anxiety

Using the extracted mobility features, we investigated whether the mobility features could predict students' SIAS scores. We studied the results of (1) the classification task that classified participants as low or high socially anxious using SIAS score=34 as a cutoff and (2) the regression task by predicting the actual SIAS score of a participant (between 0 and 80).

Table 2. Pearson correlation (with *P* values) between each participant's Social Interaction Anxiety Scale score and his or her transition frequency from one location to another. Only transitions having a correlation coefficient > 0.2 and *P*<.01 have been presented.

Transitions	Observations	Pearson coefficient	*P* value
Out of town Religious	45	0.337	0.011
Supermarket Education	46	0.339	0.010
Education Supermarket	133	0.223	0.004
Leisure Other houses	112	−0.219	0.010
Out of town Leisure	35	−0.282	0.050
Service Leisure	45	−0.291	0.026
Leisure Leisure	26	−0.495	0.005

For each day, we extracted 220 mobility features as follows: (1) cumulative staying time at each different location and during each different time window: 8 am-4 pm, 4 pm-12 am, and 12 am-8 am; (2) the distribution of visits over time (time series of locations visited during that day); (3) location entropy; (4) the frequency of each different transition; (5) the type of day (weekday or weekend); and (6) the day of the week (eg, Monday, Tuesday). Note that we did not include the actual date (ie, MM/DD/YYYY) to avoid overfitting because we were testing our approach over a short study period. In long-term studies, this feature can be added to track special events such as national holidays. Mobility patterns during these special events may be a predictor of social avoidance and social anxiety.

Theoretically, daily mobility features would all be included in the same model as predictors of trait social anxiety. However, in practice, this is not feasible because of the large number of dimensions for a small number of participants. In other words, for each day, each different place, each transition type, and each time window (morning, evening, etc) would be a different feature. Thus, the total number of features will increase with the number of days in the study. This phenomenon is known as *the curse of dimensionality* where the volume of the space increases so fast that the available data become sparse [29]. There are two traditional solutions to tackle this problem. The first is to aggregate the features on daily basis, which means, instead of having more than 3000 features for 228 participants, we determine only 220 aggregated features (average time spent at home, at university, etc during the study period). Thus, 220 is the number of distinct mobility features that we may have in a given day. We utilized this approach as one of our baseline measures (BM_1). The second solution is to apply a dimensionality reduction technique such as principal component analysis (PCA) or autoencoders to select the most important features. We also applied this method as one of our baseline measure comparisons (BM_2), reducing our feature space to between 50 and 200.

Finally, we proposed a new method to predict a daily SIAS score, which uses a neural network with all 220 mobility features and compare it to the two traditional solutions (BM_1 and BM_2). Figure 6 describes the design of our method. We started by predicting candidate SIAS scores for each participant for each day of the study (ie, if a participant had 15 days, we predicted 15 candidate SIAS scores); then, in the second layer, a global predicted SIAS score was calculated by aggregating the daily candidate SIAS scores. For the regression task, the aggregation function used a 7% trimmed mean of the predicted daily SIAS scores; the trimmed mean helps eliminate the influence of predictions on the tails that may unfairly affect the traditional mean. However, for the classification task, the dominant class was chosen. If there was no dominant class (number of days predicted as low and high were the same), the aggregation function chose one class randomly (See aggregation in Figure 6).

Our method has the following advantages:

- Predicting on daily basis allows to capture low-level mobility features that may improve the performance of our predictor (eg, time spent at some locations during the evening of Fridays) without facing the curse of dimensionality.
- After predicting a candidate SIAS score for a given day, all raw GPS data and mobility features for that day can be deleted. Consequently, such methods can operate efficiently on mobile phones, which are known for their limited storage capacity.
- The proposed model operates incrementally and is independent of the number of days. This means it can predict social anxiety levels from the first days of the study, and it can predict SIAS scores for participants having unbalanced data (fewer number of days than expected). This is not possible with other baseline methods because they require all participants to have the same number of features (dependent on number of days) to be able to train and predict social anxiety.
- By having several candidate SIAS scores per participant, we are able to deal with days that might be outliers (ie, mobility patterns significantly different). Our method is designed to remove noise (using trimmed mean) that may be generated by unusual behaviors that may bias the data. For instance, one participant may behave as low socially anxious for only one day, while his or her behavior during other days will resemble a high SAP. This outlier will be automatically removed by our model.

The extracted features are used to train a neural network. We used a multilayer perception (MLP) [30] trained using a back propagation algorithm. It uses hyperbolic tangent activation function and contains 2 hidden layers with 100 nodes each (we did not notice a better performance by adding more nodes and hidden layers). Neural networks are popular models that have

shown great promise in many tasks such as sentiment classification, image captioning, and natural language processing. We chose MLP because of its ability to detect nonlinear relationships between inputs and outputs; thus, it can detect hidden patterns between mobility and SIAS that we could not find in our linear analysis presented earlier (ie, correlations).

We compared our prediction method with two baseline models BM_1 and BM_2. BM_1 is a nonincremental method that predicts the SIAS score on biweekly basis, ie, it waits until the end of the study to predict social anxiety levels. BM_1 uses the same mobility features as our method but aggregated on the whole study period. For instance, rather than computing cumulative staying time for each location per day, BM_1 computes the average daily cumulative staying time per location for the entire study period; it also does the same for the other features: entropy of location, frequency of transitions, and average cumulative staying time during weekdays versus weekends. We also compared our method to BM_2 that uses PCA to reduce feature space from 3000 to 200 (less than the number of participants, which were 228). Note that we did not notice any improvement when we reduced feature space dimension to 50, 100, or 150. Both BM_1 and BM_2 use an MLP with a structure similar to our method (ie, 2 hidden layers, 100 nodes each).

Table 3 illustrates the overall performance of our method compared with that of BM_1 and BM_2 for both classification and regression tasks. The models were evaluated using leave-one participant-out cross-validation (LOOCV) and 10-fold cross-validation (FCV), where each validation fold contained 12 random high SAP and 12 random low SAP; the overall number of days in each fold was around 300 days. Table 3 presents three evaluation metrics averaged across folds: (1) root-mean-square error (RMSE) for the regression task and (2) accuracy and (3) F-1 measure for the classification task.

Figure 6. The architecture of the proposed prediction method. Daily mobility features are first extracted and then used to predict multiple candidate Social Interaction Anxiety Scale (SIAS) scores using a neural network. A global SIAS score is assigned to a participant by aggregating the predicted candidate SIAS scores.

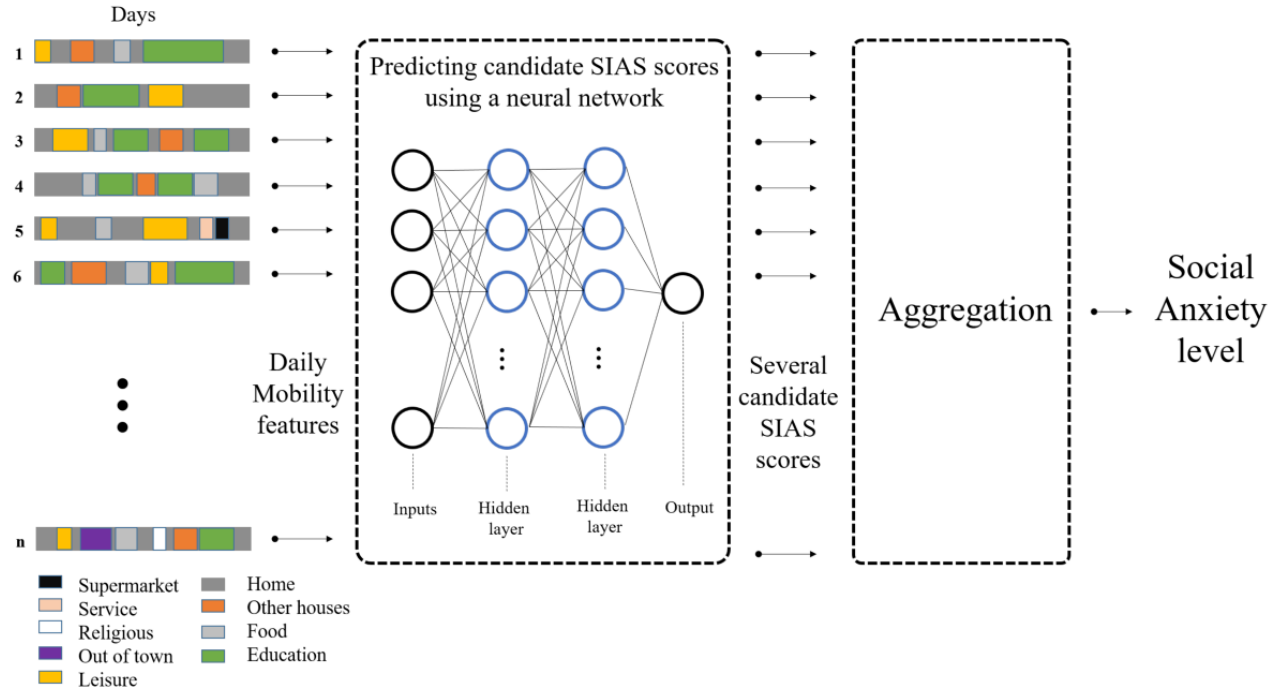
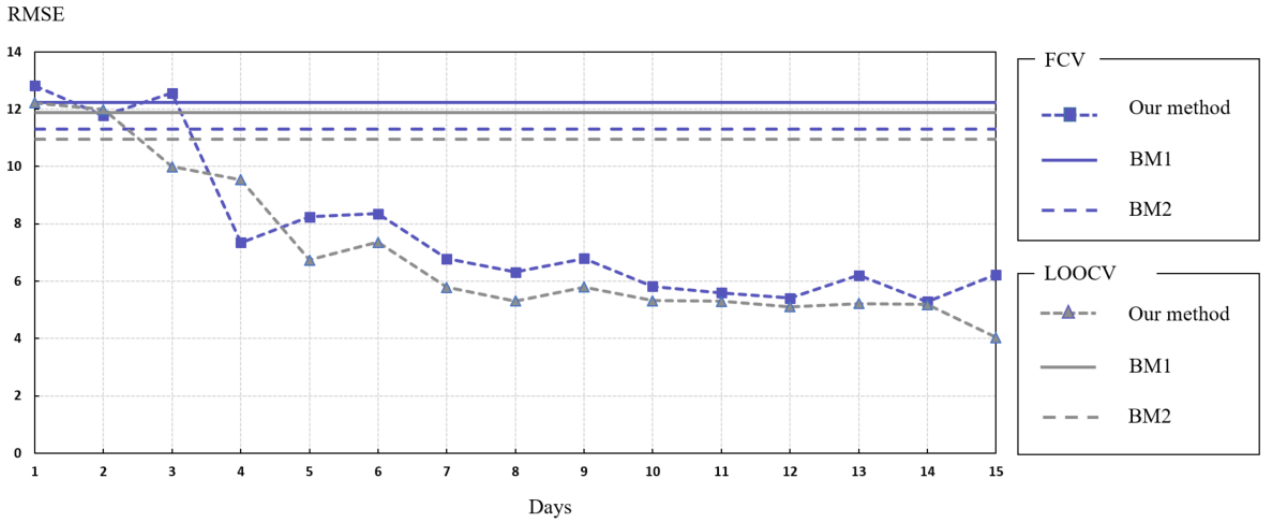


Table 3. Performance of our prediction method (OM) compared with that of two baseline methods, BM_1 and BM_2 for both classification and regression tasks. We used two evaluation methods: leave-one participant-out cross-validation (LOOCV) and 10-fold cross-validation (FCV); and three evaluation metrics: root-mean-square error (RMSE) for the regression task and accuracy and F-1 for the classification task.

Methods	LOOCV			FCV		
	RMSE	Accuracy	F-1	RMSE	Accuracy	F-1
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
OM	7.06 (2.01)	0.85 (0.04)	0.87 (0.04)	7.7 (2.95)	0.81 (0.03)	0.85 (0.04)
BM_1	11.87 (4.23)	0.69 (0.06)	0.72 (0.04)	12.23 (4.71)	0.65 (0.04)	0.69 (0.04)
BM_2	10.95 (4.92)	0.72 (0.04)	0.75 (0.04)	11.31 (4.55)	0.71 (0.04)	0.69 (0.03)

Figure 7. The performance of models over the course of the study. FCV: 10-fold cross-validation, LOOCV: leave-one participant-out cross-validation, RMSE: root-mean-square error.



Results show that our methods outperformed and for both classification and regression tasks and for both types of validation. For instance, for LOOCV, our method showed an accuracy of 85%, F-1 measure of 87%, and a RMSE of 7.06 compared with BM_1, where we recorded an accuracy of 69%, F-1 measure of 72%, and a RMSE of 11.87, and BM_2, where we recorded an accuracy of 71%, F-1 measure of 75%, and a RMSE of 10.93. The promising results of our hierarchical prediction method are justified by the fact that extracting and using daily features helps learn the different hidden patterns between mobility and SIAS that we presented above (eg, mobility in weekdays vs weekends, cumulative staying time at each different location during mornings vs evenings, and transition frequencies) and this while taking into account the day of the week; for instance, activities during Friday evenings (when students tended to engage in leisure activities) can be a predictor of social isolation. BM_1 and BM_2 failed to capture these low-level patterns that may play a role in predicting social anxiety, mainly because when aggregating on the study period basis or reducing feature space dimension, the model loses some information such as mobility during a specific time epoch (eg, morning), on a specific *type* of day (eg, Sunday).

Our prediction method operates incrementally, which means that it can predict an SIAS score starting from the first day of the study and update the estimated SIAS on each subsequent day of the study. Figure 7 depicts the average RMSE of all participants over the course of the study. The RMSE improved converging at about 7 study days. We hypothesized that this was because 1 week of data was necessary to observe behavioral patterns.

Even though our method provided superior results compared with baseline methods, it still had an error of 7.06 on an 80-point scale. One hypothesis is that this may be justified by the limitations of SIAS scores that may suffer from self-report biases [31]. Another limitation was the short study length that may not be enough to characterize the rhythms indicative of social anxiety.

Discussion

Principal Findings

In this paper, we demonstrated the feasibility of assessing college students' social anxiety through GPS-based localization. Findings from this study suggest that it is possible to use passively sensed location data from mobile phones to predict social anxiety levels. We integrated semantic labels of locations, such as locations of leisure, into our prediction models. This provided a more nuanced understanding of the behavioral patterns of socially anxious individuals. For instance, consistent with the existing theory and psychological research, we found that socially anxious students tended to avoid locations that contained a threat of social scrutiny. High SAP infrequently visited food places and engaged in less leisure activities during evenings and weekends. Exploiting the temporal richness of passively sensed location data also revealed that socially anxious students spent more time at home after school between 4 pm and 12 am. This study also extended the prior work by presenting a predictive model based on neural networks. The results of our prediction method suggest that social anxiety can be efficiently predicted by mobility features. Importantly, our findings are based on objective behavioral data gathered from people's daily lives and, therefore, avoid recall biases associated with self-reports of behavior.

While most research using mobile sensing has focused on depression, relatively little is known about the location features that might be indicative of a high social anxiety level. Social anxiety and depression share common symptoms and underlying factors, such as high negative emotionality, social withdrawal, and avoidance. However, there are also key differences. For example, depression is often marked by a slowing of movement (psychomotor retardation) and a general reduction of activity that is not found in social anxiety. Our results indicate that socially anxious individuals tend to spend time at home specifically after school, a time when additional social activities are less likely be mandatory (eg, classes are done). Thus, rather than demonstrating a general avoidance of all activities

throughout a day (which could indicate a general lack of motivation or energy), this finding sheds light on the temporal nuances of social anxiety that may not be found in depression and suggests a pattern more specific to social avoidance. There may also be key differences in the types of places that are avoided; while some individuals with high depression may avoid all locations that require high energy (eg, going to a gym), those with high social anxiety levels may focus on avoiding places and times that pose the greatest risk of social evaluation (eg, places of leisure during peak hours, which fits our results). Our findings are among the first to examine mobility features of social anxiety using fine-grained GPS data from people's daily lives. Future work leveraging mobile sensors has the potential to test and improve on existing psychological models. For example, our findings suggest that contextual factors such as location type and time of day may have an important impact on how people with social anxiety choose to isolate themselves.

Limitations

Although our findings reveal mobility features associated with social anxiety in everyday life, it is important to stress that these findings are preliminary. First, our findings are based on an unselected sample of undergraduate students. While a homogeneous sample increases the internal validity of our findings, using university students could limit the generalizability of our results to nonstudent populations whose daily life experiences and activities are different [32]. Furthermore, while there was a good representation of high socially anxious individuals in our sample who met clinical cutoffs, a nonclinical sample of undergraduate students was used. Future research should test whether the mobility features identified in this study can be generalized to a clinically diagnosed population. Second, due to the correlational nature of the data, no causal claims between social anxiety symptoms and mobility features could be made. For example, we cannot exclude the possibility that other factors outside of social anxiety influenced our findings, such as alcohol use, which is a common way of coping with distress. We also did not correct for the possible effects of multiple comparisons, given our focus of exploring mobility features that may be candidates for future research. We hope the current work will be a launching point for subsequent researchers interested in using passively sensed data for detecting social anxiety. Third, given limitations of statistical power, we were unable to examine the potential moderating roles of gender and race, which should be more closely examined in future work. Finally, one issue that future work should address is the issue of multiple comparisons, which can inflate the chance of accidental discoveries. This usually arises when the same statistical tests are repeatedly computed from the same set of observations in a hypothesis-driven research [33]. Fortunately, there are several ways of addressing this issue, many of which require researchers to set a more stringent significance threshold [34,35]. For example, while many of the transition correlations presented in the current study remained significant at $P<.01$, it is important to note that we did not have *a priori* hypotheses regarding what specific types of location pairs would be most relevant. Thus, future work should always include the direction and magnitude of effect sizes. Finally, research also suggests that different types of analyses are less prone to bias from multiple comparisons, such as those based on a hierarchical Bayesian framework [36].

Conclusion

In spite of the limitations, the ability to use fine-grained GPS data to detect behavioral patterns associated with social anxiety has many important implications. It opens up the possibility that health care professionals can identify and monitor those in need of help, but struggle with the prospect of initiating contact with others. With the increasing prevalence of social anxiety disorder and other mental health concerns, novel techniques for assessing psychological distress have become increasingly important. By leveraging the ubiquity of mobile phones and their increasingly powerful sensors, researchers and clinicians might finally be able to overcome many of the traditional obstacles to providing care.

Acknowledgments

This research was supported by the Hobby Postdoctoral and Predoctoral Fellowships in Computational Science.

Conflicts of Interest

None declared.

References

1. Stein M, Walker J. Triumph Over Shyness Conquering Social Anxiety Disorder Second Edition. New York: McGraw-Hill; 2009.
2. Brook CA, Willoughby T. The social ties that bind: social anxiety and academic achievement across the university years. J Youth Adolesc 2015 May;44(5):1139-1152. [doi: 10.1007/s10964-015-0262-8] [Medline: 25691148]
3. American College Health Association. National College Health Assessment II: Reference Group Executive Summary. Hanover, MD: American College Health Association (ACHA); 2014.
4. Kashdan TB, Collins RL. Social anxiety and the experience of positive emotion and anger in everyday life: an ecological momentary assessment approach. Anxiety Stress Coping 2010 May;23(3):259-272. [doi: 10.1080/10615800802641950] [Medline: 19326272]
5. Kashdan T, Farmer A. Differentiating emotions across contexts: Comparing adults with and without social anxiety disorder using random, social interaction, and daily experience sampling. Emotion 2010;14(3):638.

6. Boukhechba M, Bouzouane A, Bouchard B, Gouin-Vallerand C, Giroux S. Energy Optimization for Outdoor Activity Recognition. Journal of Sensors 2016;2016:1-15. [doi: 10.1155/2016/6156914]
7. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: Assessing Mental Health, Academic Performance, Behavioral Trends of College Students Using Smartphones. USA: ACM; 2014 Presented at: ACM International Joint Conference on PervasiveUbiquitous Computing; 2014; New York p. A. [doi: 10.1145/2632048.2632054]
8. Faurholt-Jepsen M, Bauer M, Kessing L. Smartphone-based objective monitoring in bipolar disorder: status and considerations. International Journal of Bipolar Disorders Internet 2018.
9. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. : . Proceedings of the 2015; 2015 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing ACM; 2015; Osaka, Japan p. 1293.
10. LiKamWa R, Liu Y, Lane N, Zhong L. MoodScope: building a mood sensor from smartphone usage patterns. : ACM Press; 2013 Presented at: MobiSys; 2013; Taiwan p. 389. [doi: 10.1145/2462456.2464449]
11. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
12. Ameko M, Cai L, Boukhechba M, Daros A, Chow P, Teachman B, et al. Cluster-based Approach to Improve Affect Recognition from Passively Sensed Data. 2017 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2018; Las Vegas p. 0029. [doi: 10.1109/bhi.2018.8333461]
13. Boukhechba M, Huang Y, Chow P, Fua K, Teachman B, Barnes L. Monitoring social anxiety from mobility and communication patterns. 2017 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2017; Hawaii. [doi: 10.1145/3123024.3125607]
14. Polzer N, Gewald H. A Structured Analysis of Smartphone Applications to Early Diagnose Alzheimer´s Disease or Dementia. Procedia Computer Science 2017;113:448-453. [doi: 10.1016/j.procs.2017.08.293]
15. Insel T. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
16. Saeb S, Zhang M, Karr C, Schueller S, Corden M, Kording K, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
17. Saeb S, Lattie E, Kording K, Mohr D. Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112 [FREE Full text] [doi: 10.2196/mhealth.7297] [Medline: 28798010]
18. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Feb 22:22. [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
19. Boukhechba M, Gong J, Kowsari K, Ameko M, Fua K, Chow P, et al. Physiological changes over the course of cognitive bias modification for social anxiety. 2018 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2018; Las Vegas. [doi: 10.1109/bhi.2018.8333458]
20. Mattick R, Clarke J. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behaviour Research and Therapy 1998 Apr;36(4):455-470. [doi: 10.1016/s0005-7967(97)10031-6]
21. Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH. Validation of the Social Interaction Anxiety Scale and the Social Phobia Scale across the anxiety disorders. Psychological Assessment 1997;9(1):21-27. [doi: 10.1037//1040-3590.9.1.21]
22. Xiong H, Huang Y, Barnes L, Gerber M. Sensus: a cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. Sensus: ACM Press; 2016 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16; 2016; Heidelberg, Germany. [doi: 10.1145/2971648.2971711]
23. Ruta M, Scioscia F, Ieva S, Loseto G, Di SE. Semantic Annotation of OpenStreetMap Points of Interest for Mobile Discovery and Navigation. 2012 Presented at: IEEE First International Conference on Mobile Services; 2012; Honolulu, HI, USA p. 39. [doi: 10.1109/mobserv.2012.17]
24. Boukhechba M, Bouzouane A, Bouchard B, Gouin-Vallerand C, Giroux S. Online recognition of people's activities from raw GPS data: semantic trajectory data analysis. 2015 Presented at: 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments Internet ACM Press; 2015; Greece. [doi: 10.1145/2769493.2769498]
25. Kang JH, Welbourne W, Stewart B, Borriello G. Extracting places from traces of locations. SIGMOBILE Mob Comput Commun Rev 2005 Jul 01;9(3):58. [doi: 10.1145/1094549.1094558]
26. Shu-Cherng F, Tsao H. -S. J. Entropy Optimization: Shannon Measure of Entropyits Properties. Boston, MA: Springer; 2011.
27. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May 08;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]

28. Rapee R, Heimberg R. A cognitive-behavioral model of anxiety in social phobia. Behaviour Research and Therapy 1997 Aug;35(8):741-756. [doi: 10.1016/s0005-7967(97)00022-3]
29. Coppejans M. Breaking the Curse of Dimensionality. In: SSRN Journal. Encyclopedia of Operations Research and Management Science Internet Dordrecht: Kluwer Academic Publishers; 2001.
30. The multilayer perceptron. In: Neural Computing: An Introduction. Bristol, United Kingdom: IOP Publishing Ltd; 1990:A.
31. Zimmerman M. Self-Report Measures. In: Encyclopedia of Clinical Neuropsychology. New York, NY: Springer New York; 2011.
32. Peterson R. On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. Journal of Consumer Research 2001.
33. Ryan T. Multiple comparison in psychological research. Psychological Bulletin 1959;56(1):26-47.
34. McLaughlin M, Sainani K. Bonferroni, Holm, and Hochberg corrections: fun names, serious changes to p values. PM R 2014 Jun;6(6):544-546. [doi: 10.1016/j.pmrj.2014.04.006] [Medline: 24769263]
35. Multiple Comparisons. In: Chapman & Hall/CRC Biostatistics Series. United Kingdom: Chapman and Hall/CRC; 2011:167-180.
36. Gelman A, Hill J, Yajima M. Why We (Usually) Don't Have to Worry About Multiple Comparisons. Journal of Research on Educational Effectiveness 2012.

Abbreviations

FCV: 10-fold cross-validation
GPS: Global Positioning System
LOOCV: leave-one participant-out cross-validation
MLP: multilayer perception
OSM: open street map
RMSE: root-mean-square error
SAP: socially anxious participant
SIAS: Social Interaction Anxiety Scale

Edited by J Torous, R Calvo, M Czerwinski, G Wadley; submitted 12.02.18; peer-reviewed by S Schueller, M Larsen, I Barnett; comments to author 13.03.18; revised version received 20.04.18; accepted 16.05.18; published 04.07.18.

Please cite as:
Boukhechba M, Chow P, Fua K, Teachman BA, Barnes LE
Predicting Social Anxiety From Global Positioning System Traces of College Students: Feasibility Study
JMIR Ment Health 2018;5(3):e10101
URL: http://mental.jmir.org/2018/3/e10101/
doi:10.2196/10101
PMID:29973337

Review

Digital Characteristics and Dissemination Indicators to Optimize Delivery of Internet-Supported Mindfulness-Based Interventions for People With a Chronic Condition: Systematic Review

Lahiru Russell[1], MSc (Biochem), PGDip (Epi); Anna Ugalde[1], BA (Hons), PhD; Donna Milne[2,3], RN, PhD; David Austin[4], Grad Dip Psych, Grad Dip Appl Sci, PhD; Patricia M Livingston[1], BA (Hons), PhD

[1]School of Nursing and Midwifery, Deakin University, Geelong, Australia

[2]Melanoma and Skin Services, Peter MacCallum Cancer Center, Melbourne, Australia

[3]Department of Cancer Experiences Research, Peter MacCallum Cancer Center, Melbourne, Australia

[4]School of Psychology, Deakin University, Geelong, Australia

Corresponding Author:
Lahiru Russell, MSc (Biochem), PGDip (Epi)
School of Nursing and Midwifery
Deakin University
11 Gheringhap Street
Geelong, 3220
Australia
Phone: 61 392446912
Email: l.russell@deakin.edu.au

Abstract

Background: Internet-supported mindfulness-based interventions (MBIs) are increasingly being used to support people with a chronic condition. Characteristics of MBIs vary greatly in their mode of delivery, communication patterns, level of facilitator involvement, intervention period, and resource intensity, making it difficult to compare how individual digital features may optimize intervention adherence and outcomes.

Objective: The aims of this review were to (1) provide a description of digital characteristics of internet-supported MBIs and examine how these relate to evidence for efficacy and adherence to the intervention and (2) gain insights into the type of information available to inform translation of internet-supported MBIs to applied settings.

Methods: MEDLINE Complete, PsycINFO, and CINAHL databases were searched for studies assessing an MBI delivered or accessed via the internet and engaging participants in daily mindfulness-based activities such as mindfulness meditations and informal mindfulness practices. Only studies using a comparison group of alternative interventions (active compactor), usual care, or wait-list were included. Given the broad definition of chronic conditions, specific conditions were not included in the original search to maximize results. The search resulted in 958 articles, from which 11 articles describing 10 interventions met the inclusion criteria.

Results: Internet-supported MBIs were more effective than usual care or wait-list groups, and self-guided interventions were as effective as facilitator-guided interventions. Findings were informed mainly by female participants. Adherence to interventions was inconsistently defined and prevented robust comparison between studies. Reporting of factors associated with intervention dissemination, such as population representativeness, program adoption and maintenance, and costs, was rare.

Conclusions: More comprehensive descriptions of digital characteristics need to be reported to further our understanding of features that may influence engagement and behavior change and to improve the reproducibility of MBIs. Gender differences in determinants and patterns of health behavior should be taken into account at the intervention design stage to accommodate male and female preferences. Future research could compare MBIs with established evidence-based therapies to identify the population groups that would benefit most from internet-supported programs.

Trial Registration: PROSPERO CRD42017078665; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=78665 (Archived by WebCite at http://www.webcitation.org/71ountJpu)

(JMIR Ment Health 2018;5(3):e53) doi:10.2196/mental.9645

KEYWORDS

mindfulness; internet; chronic condition

Introduction

Background

Over the past two decades, mindfulness has become a major focus of research in health care settings for people with chronic conditions [1]. In clinical research, the application of mindfulness focuses on cognitive and emotional regulation to help patients cope better with their conditions [2]. There is evidence to support the use of mindfulness-based interventions (MBIs) such as mindfulness-based stress reduction (MBSR) and mindfulness-based cognitive therapy (MBCT) for mental and physical symptoms in people with cancer, cardiovascular disease, chronic pain, depression, and anxiety disorders [1].

The internet has become a tool to disseminate MBIs to a larger number of people, bypassing barriers related to cost and time commitments for in-person therapy [3], the need for a trained therapist [4], and reluctance to engage in group interventions [5]. Recently, a meta-analysis and a systematic review assessed the impact of internet-supported MBIs on mental health [6] and people with chronic physical conditions [7]. The studies showed promising results for improving mental health outcomes and alleviating symptom burden. The meta-analysis was conducted among a diverse group of people (with or without a chronic illness) and reported beneficial small-to-moderate effects of the interventions on depression, anxiety, stress, well-being, and mindfulness [6]. Nevertheless, as highlighted by the authors, interventions varied in their modes of delivery (mobile phone app, website, or Web-based classroom), therapeutic approaches (mindfulness-based vs mindfulness-inspired interventions), and level of therapist involvement (self-guided vs therapist-guided) [6]. This makes it difficult to determine which aspects of the Web-based designs optimized intervention delivery and their associated behavior change. Another systematic review involving people with physical conditions showed a positive effect of the intervention compared with usual care on a variety of outcomes such as pain acceptance, coping measures, and depressive symptoms [7]; however, results were mixed when the interventions were compared with an active control group, such as cognitive behavioral therapy. The intervention delivery mode was broadly categorized into synchronous (ie, real-time delivery such as instant messaging, telephone, or videoconferencing) versus asynchronous (such as emails) and facilitated (therapist or moderator involvement) versus self-guided. Interventions can therefore vary across a wide range of digital features used for various purposes. For example, sending reminders and providing personalized feedback through emails are both asynchronous functions but may influence intervention engagement differently [8,9]. It is, therefore, important to examine the technology used in internet-supported MBIs to understand how digital functions optimize intervention delivery and outcomes.

The challenges of translating research findings into practice are well documented [10,11]. The lack of measures assessing generalizability and sustainability of interventions in trials is a critical factor hindering translation of findings [10]. Pragmatic frameworks used for program implementation and outcome evaluation can help bridge the gap between scientific knowledge and dissemination [12]. These frameworks tend to combine factors ascertaining the internal validity of a program, such as changes in outcomes of interest and attrition and adherence rates, with concepts relevant to external validity, such as representativeness of study population, availability and cost of resources, and organizational readiness [12], which may have particular relevance for Web-based mindfulness interventions. Hence, assessing the efficacy of internet-supported MBIs while collecting information relevant to its generalizability will provide important information on the potential impact on wider communities [13].

Aims

The aims of this review were to (1) provide a description of digital characteristics of internet-supported MBIs and examine how these relate to evidence for efficacy and adherence to the intervention and (2) gain insights into the type of information available to inform translation of internet-supported MBIs to applied settings.

Methods

Review Process

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines [14]. Due to the heterogeneity of the study designs, populations, and outcomes, a narrative synthesis of the results was conducted rather than a meta-analysis. The protocol was registered on PROSPERO database on 01/11/2017 with reference number CRD42017078665.

Eligibility Criteria

The review focused on internet-delivered MBIs for people with a chronic condition. Inclusion criteria were structured according to the PICOS framework [15]. The PICOS acronym stands for patient (P); intervention (I); comparison, control, or comparator (C); outcome (O); and study type (S) and is described in more detail below.

Participants

Participants were aged 18 years or older and diagnosed with a chronic condition such as, but not limited to, heart disease, diabetes, cancer, respiratory disease, or mental illness (eg, depression).

Interventions

Interventions were MBIs that met the following two criteria: (1) delivered or accessed via the internet with at least 50% of interactions being technology-mediated and (2) engaging participants in daily mindfulness-based activities such as mindfulness meditations and informal mindfulness practices. Studies were excluded if they examined mindfulness as a component of another treatment such as acceptance and

commitment therapy [16] and dialectical behavior therapy [17,18], as it was not possible to dissociate the effect of mindfulness from other components of the intervention.

Control Group

Control group could be comparison groups of alternative interventions (active comparator), usual care, or wait-list.

Outcome Measures

All outcome measures were considered.

Study Type

This review included original papers reporting on randomized, quasi-experimental, and feasibility or pilot studies comparing the efficacy of a Web-based MBI with a control group. Cross-sectional studies, case reports, review articles, dissertations, and commentaries were excluded from this review.

Study Selection

Web-based psychoeducation studies were first evident in the literature in 2000 [19]. Hence, the literature search for this review was conducted between January 2000 and July 2017 across three Web-based databases (MEDLINE Complete, PsycINFO, and CINAHL) using the following search terms: Online (online OR internet OR "web-base" OR ehealth OR etherap* OR app* OR telehealth OR telemedicine), Mindfulness (mindful* OR MBSR OR MBCT OR "acceptance and commitment therapy" OR awareness OR meditat*), and Intervention (intervention* OR therap* OR group* OR treatment*). Identified papers and key review papers [6-8] were further examined for additional eligible studies.

Given the broad definition of chronic conditions (ie, long lasting with persistent effects [20], in which conditions may deteriorate, advance, fluctuate, or be characterized by remissions [21]), specific conditions were not included in the original search to maximize results.

Papers published in English that met the eligibility criteria were included in the review. Additionally, for any paper meeting the eligibility criteria, data were extracted from related papers describing different aspects of the same study (eg, methods paper and cost-effectiveness paper).

Review Process

Titles of identified records were screened by one author (LR). Papers not meeting eligibility criteria were excluded at this stage, and abstracts of remaining papers were read by two authors (LR and DM). Full texts of abstracts meeting eligibility criteria were reviewed, and data were extracted by LR. In case of ambiguity, studies were discussed and agreed upon with coauthor AU.

Data Extraction

Study Characteristics

A standardized data extraction form was developed to collect information about study design, assessment time points, primary outcomes measures, participant characteristics, intervention and control conditions, intervention adherence, study findings, and attrition rates.

Study findings were categorized into whether the intervention group had a statistically significant improvement (yes or no) on the primary outcomes at the postintervention assessment compared with the control group.

Digital Features of Internet-Supported Mindfulness-Based Interventions

Reporting on digital features was guided by a coding scheme developed by Webb and colleagues for Web-based interventions [22], and the features described in the studies have been included in this review. These features were divided into six main categories: (1) delivery mode, (2) navigational format, (3) automated communication, (4) additional material (eg, ebook, video, or audio files), (5) other features (eg, book or hard copy of intervention), and (6) level of facilitator involvement. On the basis of the information reported in the included studies, the delivery mode was further divided into Web-based, videoconference, and email-based. Navigational format was defined as tunneled (the intervention could only be experienced in a predetermined order, and modules, sessions, or Web pages could not be skipped) or flexible (the content of the intervention could be accessed according to the user's preference, and modules, sessions, or Web pages could be skipped) [23]. Automated communication was divided into email reminders and follow-up messages to encourage participation. Facilitator involvement can vary substantially across interventions [8,24,25]; therefore, the level of involvement was summarized using criteria similar to those of another review [8]: interventions without any facilitator involvement were categorized as *none (self-guided)*; interventions where facilitators were only providing reminders, links to modules, encouragement, and answering logistical questions were categorized as having *low facilitator involvement*; *medium facilitator involvement* referred to the provision of feedback on homework for mastering mindfulness skills; and *high level of facilitator involvement* referred to the provision of intervention in person.

Internal and External Validity Indicators

Glasgow and colleagues developed a framework to evaluate the degree to which behavioral interventions reported on efficacy (internal validity) and generalizability to other settings and populations (external validity) [26]. More specifically, the framework focuses on the reporting of the following five dimensions: (1) the *reach* into the target population and representativeness of the study sample; (2) *efficacy or effectiveness* of the intervention on primary outcome(s) tested under either restricted or controlled or real-world conditions, quality of life, and avoidance of unintended or negative consequences; (3) *adoption* rates of organizations and staff that would use the intervention and the characteristics of those organizations and staff; (4) *implementation* of the intervention as intended; and (5) *maintenance* of the effects at the individual level and sustainability of the intervention at an organizational or delivery level (RE-AIM: reach, efficacy/effectiveness, adoption, implementation, maintenance). The RE-AIM framework has been used to review the literature in diverse health areas, such as physical activity during pregnancy [27] or among family caregivers [28], self-management programs for diabetes [29,30], and health literacy interventions [31].

The degree to which internal and external validity were reported was recorded using a 21-item validated data extraction tool capturing the five dimensions of the RE-AIM framework [32]. Each dimension comprises specific indicators that were rated as criteria met (*yes*) or not met (*no*) and had equal weight. Each indicator reported was given a score of 1 (see Multimedia Appendix 1).

Reach

The following information addressing the internal validity of each study was coded: methods used to identify the target population and its inclusion and exclusion criteria. The following indicators addressed external validity: the number of individuals who agreed to participate compared with the total number of eligible participants (participation rate) and the characteristics of participants compared with nonparticipants (representativeness).

Efficacy/Effectiveness

Efficacy studies investigate the effects of an intervention under highly controlled conditions with a homogenous patient population enrolled using strict inclusion and exclusion criteria [33]. Effectiveness or pragmatic studies examine interventions under conditions similar to real-world practice, such as routine clinical settings, with more heterogeneous patient populations. Effectiveness studies may also use a randomized controlled trial (RCT) design; however, the intervention is more often compared with usual care [33]. The efficacy/effectiveness dimension is also composed of indicators strongly associated with internal validity such as changes in primary outcomes and the proportion of participants lost at follow-up (attrition rate). Other indicators include the type of analysis conducted (ie, intention-to-treat or completer analysis) and measures of quality of life. We also examined if papers assessed changes in mindfulness scores and proposed a potential mechanism of action for mediation or moderation effects of these scores on the intervention [34]. For example, did the strength of the relationship between the intervention and the outcome vary according to participants' mindfulness scores (moderation effect of mindfulness)? Or, does the intervention cause changes in mindfulness scores, which in turn impact the outcome measures (mediation effect)?

Adoption

This dimension assessed the extent to which an intervention is carried out at a staff and setting level. Papers were reviewed to identify the degree to which intervention settings were described (eg, primary care, outpatient clinics, and online forums). Additionally, methods to identify the staff who delivered the intervention and their level of expertise were also coded.

Implementation

The duration and frequency of the intervention, the extent to which the protocol was delivered as intended (adherence rate), and the cost of delivery were coded as indicators for the implementation dimension.

Maintenance

This dimension, also a measure of sustainability, was coded for indicators reporting on assessments 6 or more months after the completion of the intervention, the level of maintenance of the intervention, and the cost associated with this maintenance.

The quality of reporting on RE-AIM indicators was calculated for each study with a possible score ranging from 0 to 21. Following criteria from previous RE-AIM reviews [31,35], the reporting quality was categorized as high, moderate, or low for studies scoring 15-21, 8-14, or less than 8, respectively.

Results

Review Process

A flow diagram of the selection process of the paper is provided in Figure 1.

The electronic database and external reference list searches produced 691 records after removal of duplicates. Title screening excluded 643 records leaving 48 abstracts that were examined, with 14 selected for full review. Two papers reported on findings from the same trial, one reported on the efficacy of the intervention to improve physical and psychological outcomes postintervention [36] and the other reported the 12-month follow-up assessment [37]. Both papers were included, but the methodology and findings were presented as one study. Another paper was a secondary analysis exploring the association of age, sex, and cancer stage on patient-reported outcomes postintervention [38]. This analysis did not include a comparative group and was therefore excluded from the review. Two other studies met the inclusion criteria, but their interventions combined two delivery modes (Web-based and telephone-based), and outcome data were not reported by delivery mode [39,40]. As effect size, attrition, and adherence rates for the Web-based group were not available, these latter studies were also excluded.

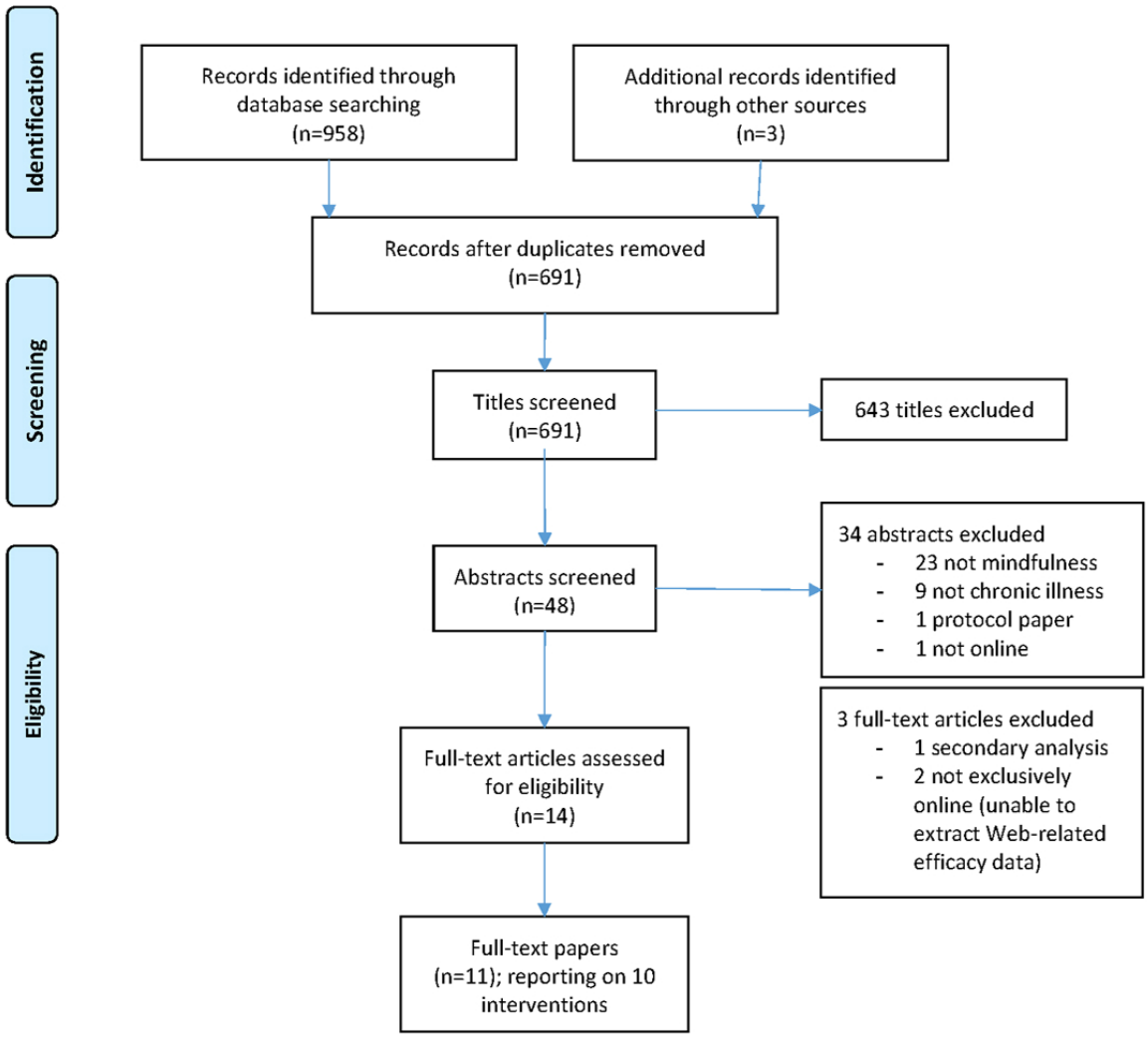
In total, 11 papers reporting 10 studies were included in the analysis.

Study Characteristics

Table 1 provides a description of the study design, Table 2 provides a description of the participant characteristics, and Table 3 provides a description of the intervention and control conditions and a summary of intervention adherence, attrition rates, and overall outcome for each study.

Studies were published from 2008 and conducted in Sweden (n=3) [43,44,49], Canada (n=2) [41,45], the United States (n=2) [42,46], the Netherlands (n=1) [36], Ireland (n=1) [47], and Germany (n=1) [48]. Out of these, 8 studies were RCTs [36,42-45,47-49] and two were quasi-experimental designs [41,46]. Most studies involved female participants with an overall mean of 74.8%, ranging between 46.3% (150/324) [36] and 98% (77/98) [42]. The overall mean age of participants was 45 years (mean age range: 36-57.6 years). Three studies comprised active comparison groups, including the same MBI delivered in person [41], a Web-based behavioral activation condition [44], and a progressive muscle relaxation program [48]. Control conditions included attention control (n=1) [42], online discussion forum (n=2) [43,49], psychoeducational program (n=1) [47], wait-list (n=2) [41,45], or usual care (n=2) [36,46]. One study involved two comparison groups [41].

Figure 1. Flow diagram of the paper selection process.



Chronic conditions examined in studies included chronic pain (n=3) [41,47,49], fibromyalgia (n=1) [42], heart disease (n=1) [36], cancer post-treatment (n=1) [45], anxiety disorder (n=1) [43], major depressive disorder (n=1) [44], residual depressive symptoms (n=1) [46], and psychosis (n=1) [48].

Primary outcome measures assessed pain or pain-related concepts (eg, pain interference, pain catastrophizing, pain-coping efficacy, and pain acceptance) in 4 studies [41,42,47,49]; depression in 3 studies [44,46,48]; distress in 2 studies [47,49]; and exercise capacity [36], anxiety [43], and mood disturbance [45] in one study each. In addition, 5 studies did not distinguish a primary outcome measure from their overall measures [41,42,47-49], which included constructs related to stress, affect, and quality of life. Study duration ranged from 6 weeks [42,48] to 12 months [36], with half of the studies including follow-up assessments [36,43,44,46,47] that ranged from 10 weeks [46] to 12 months [36]. Four studies assessed the participants' mindfulness scores [45-47,49].

Duration of the interventions ranged from 6 [42,47,48] to 10 weeks [41], with half of the interventions conducted for 8 weeks [43-46,49]. Most studies demonstrated a benefit of the internet-supported MBI compared with a control condition [36,41-43,45,46,49]. However, in one study where the intervention was compared with two different conditions (wait-list and face-to-face MBI) [41], the intervention and the face-to-face MBI groups showed improvements in mental health–related quality of life, pain catastrophizing, and usual level of pain when compared with the wait-list group. No significant difference was observed between the intervention and face-to-face MBI groups. Similarly, 3 other studies [44,47,48] did not observe any difference between the intervention and control conditions, which were either active [44,48] or psychoeducational [47] conditions. In one study, a self-help manual about progressive muscle relaxation with audio files was emailed to the participants. The program showed to be as effective at improving depressive and obsessive-compulsive symptoms in people with psychosis as the MBI [48]. In another study, a pain management psychoeducational program and an MBI, comparable in delivery mode, structure, and time commitment, were both effective in improving pain-related outcomes and subjective well-being in people living with chronic pain [47]. In the third study, a subgroup analysis found that the intervention, a Web-based MBCT, worked better than the control condition among participants with milder depression, whereas the control condition, a behavioral activation program, had a greater effect on people with severe depression [44].

Adherence to the intervention was reported in 9 of the 10 studies, but the measure of adherence varied, from objective measures, such as the number of diaries or questionnaires completed, the number of Web-based *clicks*, or videoconference sessions attended, to participant self-report. In addition, adherence rate description was also varied, reported as either the mean proportion of sessions completed, the proportion of participants who completed all sessions or viewed all pages, or those who completed at least half of the program. Table 4 describes how intervention adherence was measured and defined in each study.

Attrition rates at postintervention follow-up ranged from 11% [43] to 62% [46] in the intervention groups, whereas attrition rates among the control conditions ranged between 0% [45,46] and 49% [41].

Digital Features of the Interventions

Table 5 presents the main digital characteristics used in each intervention. Eight of the ten studies maintained the structure of in-person MBIs from which they were derived (ie, suggested a sequence of sessions or modules and recommendations for daily practice) [36,41-43,45-47,49]. However, the format was modified to match the way the internet-delivered interventions were typically applied, which resulted in briefer sessions and shorter meditation practices. The following section describes the digital features of each intervention based on their mode of delivery.

Table 1. Study design.

Reference	Study design	Time points	Primary outcome measures
Gardner-Nix et al, 2008 [41]	QE[a]	Pre, post	Physical and mental quality of life, pain catastrophizing, usual pain level, and pain-related suffering
Davis and Zautra, 2013 [42]	RCT[b]	Pre, daily for 6 weeks	Pain and pain coping efficacy, positive and negative affect, social activity engagement, loneliness, family stress, stress coping efficacy, and family enjoyment
Boettcher et al, 2014 [43]	RCT	Pre, post, 6-month follow-up	Anxiety
Ly et al, 2014 [44]	RCT	Pre, post, 6-month follow-up	Depression
Zernicke et al, 2014 [45]	RCT	Pre, post	Mood disturbance[c]
Dimidjian et al, 2014 [46]	QE	Pre, post, 10-weeks follow-up (FU1), 6-month follow-up (FU2)	Depression
Younge et al, 2015 [36]; Gotink et al, 2017 [37]	RCT	Pre, post, 12-month follow-up	Exercise capacity
Dowd et al, 2015 [47]	RCT	Pre, post, 6-month follow-up	Pain interference and distress
Moritz et al, 2015 [48]	RCT	Pre, post	Paranoia, obsessive-compulsive disorder, depression, and psychic experience scale
Henriksson et al, 2016 [49]	RCT	Pre, post	Pain intensity, pain acceptance, interference or suffering caused by pain, mindfulness, affective distress, life satisfaction, and life control

[a]QE: quasi-experimental.

[b]RCT: randomized controlled trial.

[c]Primary outcome was feasibility, but sample size was calculated with adequate power to reducing mood disturbance (secondary outcome).

Table 2. Participant characteristics.

Reference	Participant condition	Country	Females, n (%)	Age, mean (SD or range)
Gardner-Nix et al, 2008 [41]	Chronic pain	Canada	162 (75.3)	52 (32-79)
Davis and Zautra, 2013 [42]	Fibromyalgia	United States	77 (98)	46.14 (22-81)
Boettcher et al, 2014 [43]	Anxiety disorders	Sweden	65 (71)	38 (10.3)
Ly et al, 2014 [44]	Major depressive disorder	Sweden	57 (70)	36 (10.8)
Zernicke et al, 2014 [45]	Cancer recovery	Canada	45 (73)	57.6 (10.8)
Dimidjian et al, 2014 [46]	Residual depressive symptoms	United States	146 (73.0)	47.4 (11.43)
Younge et al, 2015 [36]; Gotink et al, 2017 [37]	Heart disease	The Netherlands	150 (46.3)	43.2 (27.5)
Dowd et al, 2015 [47]	Chronic pain	Ireland	112 (90.3)	44.5 (12.3)
Moritz et al, 2015 [48]	Psychosis	Germany	52 (58)	37.7 (9.7)
Henriksson et al, 2016 [49]	Chronic pain	Sweden	100 (93.5)	51 (9.3)

Table 3. Description of intervention and control conditions and summary of outcomes.

Reference	Control description	Intervention		Outcomes		
		Type	# of sessions, duration in weeks (intention)	Adherence to intervention, n (%)	Intervention group improvement over controls	Attrition (%)
Gardner-Nix et al, 2008 [41]	Active control on-site (C1) wait-list (C2)	MBCPM[a]	10 sessions per 10 weeks	NR[b]	Yes (except for physical quality of life and pain-related suffering)[c]	I[d]: 30; C1: 49; C2: 10
Davis and Zautra, 2013 [42]	Active control (health tips)	MSER[e]	12 modules per 6 weeks	Completed all modules, 19 (49)	Yes (except for pain and negative affect)	I: 15; C[f]: 5
Boettcher et al, 2014 [43]	Active control (discussion forum)	MBI[g]	16 hours per 8 weeks	All exercises completed on average, 46%	Yes	I: 11; FU[h]: 22; C: 4
Ly et al, 2014 [44]	Active control (behavioral activation)	MBCT[i]-inspired	8 weeks	Full adherence, 32 (78)	No, but after subgroup analysis: I>C for mild depression; C>I for severe depression	I: 12; FU: 17; C: 10; FU: 12.5
Zernicke et al, 2014 [45]	Wait-list	MBCR[j]	8 sessions per 8 weeks + 6 hours online silent retreat	Completed at least half the program (including retreat), 25 (83)	Yes	I: 27; C: 0
Dimidjian et al, 2014 [46]	Usual care	MBCT-inspired	8 sessions per 8 weeks	Completed all 8 sessions, 42 (42)	Yes	I: 62; FU1: 65; FU2: 73; C: 0
Younge et al, 2015 [36]; Gotink et al, 2017 [37]	Usual care	MBSR[k]-inspired	12 weeks	Completed at least half of the program, 115 (53.5)	Yes	I: 22; FU: 26; C: 16.5; FU: 22
Dowd et al, 2015 [47]	Psychoeducational	MBCT-inspired	12 sessions per 6 weeks	Viewed all sessions, 17 (74)	No	I: 55; FU: 63; C: 40; FU: 56
Moritz et al, 2015 [48]	Active control (progressive muscle relaxation)	MBI	6 weeks	Fully read the manual, 23 (61)	No	I: 26; C: 31
Henriksson et al, 2016 [49]	Active control (online forum)	MBSR-inspired	16 hours per 8 weeks	Completed full program, 18 (50)	Yes	I: 35; C: 21

[a]MBCPM: mindfulness-based chronic pain management.

[b]NR: not reported.

[c]Greater than wait-list, but not greater than onsite comparison group.

[d]I: intervention.

[e]MSER: mindful socioemotional regulation.

[f]C: control.

[g]MBI: mindfulness-based intervention.

[h]FU: follow-up.

[i]MBCT: mindfulness-based cognitive therapy.

[j]MBCR: mindfulness-based cancer recovery.

[k]MBSR: mindfulness-based stress reduction.

Table 4. Study definitions of intervention adherence and adherence rates.

Reference	Adherence defined as	Adherence rate
Gardner-Nix et al, 2008 [41]	Not reported	Not reported
Davis and Zautra, 2013 [42]	Number of diaries completed[a]	49% completed all modules
Boettcher et al, 2014 [43]	Number of *clicks*	46% of the mindfulness exercises completed
Dimidjian et al, 2014 [46]	Self-report	42% completed all sessions
Ly et al, 2014 [44]	At least one reflection emailed per week	78% completed all sessions
Zernicke et al, 2014 [45]	Number of videoconference sessions attended	83% completed at least half the program
Dowd et al, 2015 [47]	Self-report	74% viewed all sessions
Moritz et al, 2015 [48]	Self-report	61.5% fully read the manual
Younge et al, 2015 [36]	Number of assignments completed	53% completed at least half of the program
Henriksson et al, 2016 [49]	Self-report and verified by user-logged data	50% completed the full program

[a]Payment incentives for completing each diary.

Table 5. Digital features of internet-delivered mindfulness-based interventions.

Reference	Delivery mode	Navigational format	Automated communication	Additional material	Nondigital features	Level of facilitator involvement
Gardner-Nix et al, 2008 [41]	Videoconference	Tunneled	N/A[a]	CD	N/A	High
Davis and Zautra, 2013 [42]	Email-based	Tunneled	N/A	Animations and audios	N/A	Low
Boettcher et al, 2014 [43]	Web-based	Tunneled	Follow-up email	Video and audios	N/A	None (self-guided)
Dimidjian et al, 2014 [46]	Web-based	Flexible	N/A	Videos and audios	N/A	None (self-guided)
Ly et al, 2014 [44]	Web-based[b]	Flexible	N/A	Audios	N/A	Medium
Zernicke et al, 2014 [45]	Videoconference	Tunneled	N/A	Videos, audios, headsets, and webcam	Program manual	High
Dowd et al, 2015 [47]	Web-based	Flexible	Email reminders	Videos and audios	N/A	None (self-guided)
Moritz et al, 2015 [48]	Email-based	Flexible	N/A	Intervention manual (PDF) and audio files	N/A	None (self-guided)
Younge et al, 2015 [36]	Web-based	Flexible	Email reminders + follow-up text message	Videos and audios	Mindfulness book	None (self-guided)
Henriksson et al, 2016 [49]	Web-based	Flexible	Email reminders + follow-ups	Videos and audios	N/A	None (self-guided)

[a]N/A: not applicable.

[b]The intervention was delivered through a mobile app for iPhone owners or through a mobile phone−based app for other mobile phones.

Web-Based Interventions

Web-based interventions were the most common mode of delivery with six interventions out of ten being accessible through websites [36,43,44,46,47,49]. One of these interventions was described as a mobile phone–based app [44], where the intervention was accessible through participants' mobile phones. All six interventions offered meditation audio files and five offered a flexible navigational format [36,44,46,47,49]. Email reminders and follow-up messages were common features of these interventions, with four studies using automated email functionality [36,43,47,49] and one using therapist-initiated email [44]. Five of the 6 Web-based interventions were self-guided.

Email-Based Interventions

Two studies delivered the intervention via email, with one allowing for a flexible navigational format [48] and the other using a tunneled format [42]. In the former, participants were emailed a link to download a 15-page manual and four audio files providing instructions for meditation tasks. The study was self-guided, and there was no interaction with participants for

the duration of the study period (6 weeks). In the latter study, participants were emailed one module of the intervention at a time, following completion of a diary. The material in each module was delivered via Adobe Presenter, which allowed visual presentation of texts and animated pictures that accompanied an audio recording of module content. Participants were also provided with audio recordings of mindfulness meditations and were encouraged to access the meditation daily.

Videoconference Intervention

Two studies offered MBI through a videoconferencing mode [41,45], which allowed for a synchronous delivery of the intervention and most closely resembled in-person formats.

Both programs consisted of weekly 2-hour sessions and the provision of meditation audio files. However, in one study [41], the intervention took place at the participant's local hospital, whereas in the other study [45], participants accessed the intervention from their home through a Web-based educational platform that simulated a virtual classroom, where participants could see, hear, and interact in real time with other group members and the instructor. These two studies required a high level of facilitator involvement.

Internal and External Validity Indicators

Table 6 provides the proportion of internet-supported MBI studies reporting on RE-AIM dimensions and indicators.

Reach

Reach was the second most reported dimension at 66%. Studies consistently reported on the methodology for recruiting participants. Some studies recruited participants from known target populations, such as medical records [46], population registries [45,47], or outpatient clinics [36,41,49], and others employed a convenience sampling approach through the use of media outreach with Web-based and/or newspaper advertisements [42-45,48].

Inclusion and exclusion criteria were also regularly reported with only one study providing minimal description [41]. Two studies targeted mindfulness meditation–naïve participants and specifically excluded individuals with previous experience [43,45]. All studies reported on sample size, which ranged from 53 to 324 with a median of 99. Only 3 studies out of 10 provided information about participation rates [36,45,47], which were 31%, 36%, and 10%, respectively. None of the studies indicated the degree to which study samples were representative of a wider population; however, one study compared the baseline scores of the study sample on the mental component of a health-related quality of life questionnaire (SF-36v2) with the national population [41]. The mental health components comprised vitality, role emotional, social functioning, and mental health domains. These scores were 1.5 to 2 SDs below the average values of the national (US) population.

Efficacy/Effectiveness

With an overall of 75% indicators reported, efficacy or effectiveness was the most reported dimension across the studies. Changes in primary outcomes and attrition were described in all studies, but only half reported results of at least one follow-up [36,43,44,46,47], with 6-month follow-up being most common. Intention-to-treat analysis was used by majority of the studies (9 out of 10), one study reported on present-at-follow-up data [41], and one study reported on both intention-to-treat and completer analysis [36].

Most studies (n=7) reported on efficacy or effectiveness, with two [43,45] reporting on efficacy and five on effectiveness [36,41,42,48,49]. Six studies reported on quality of life or potential negative outcomes [36,41,43,44,47,49]. Five studies reported on interventions that improved participants' quality of life [41,43,44,47,49], and one found no effect [36]. No negative outcomes were reported.

Only four studies examined changes in mindfulness as a result of their interventions. Of those, three studies reported an improvement in participants' mindfulness scores [45,46,49], whereas one study reported a decrease [47].

Adoption

Adoption was the least reported dimension at 12%. Two studies described the staff who delivered the intervention—a trained research assistant [42], a study investigator, and a medical secretary [36]. Two other studies provided the level of expertise of the staff who delivered the intervention—a final year masters-level psychology student [44] and a clinician specialized in behavioral medicine with 15 years of experience in teaching MBSR [45]. Studies did not report on the identification of staff who delivered the intervention, inclusion and exclusion criteria of the delivery agent, or the adoption rate of the delivery agent.

Three studies described the intervention location—local hospitals [41], an outpatient clinic [36], and a cancer center [45]. Inclusion and exclusion criteria and adoption rate of settings were not reported.

Implementation

The mean proportion of reporting on implementation indicators was 63%. Intervention duration and frequency were reported by all studies. Nine out of ten studies reported on the extent to which the protocol was delivered as intended, but the cost of implementation was not reported in any study. Two studies offered monetary incentives to intervention participants by way of a gift voucher at enrolment [46] or a payment per returned diary [42].

Maintenance

Maintenance was the second least reported dimension at 17%. Half of the studies reported on outcome assessments at 6 months following the intervention.

Program-level maintenance and its associated costs were not reported in any study.

Overall Quality of Reporting on RE-AIM Indicators

The average reporting score was 9.4 out of 21, with scores ranging from 7 to 13. Two studies had low reporting quality, both with a score of 7 [41,48], and the other studies had moderate reporting quality with scores ranging from 8 [42,49] to 13 [36]. No study had a high reporting quality.

Table 6. Proportion of internet-delivered mindfulness-based intervention studies reporting on RE-AIM (reach, efficacy/effectiveness, adoption, implementation, maintenance) dimensions and indicators (N=10).

Indicator		Studies reporting, n (%)
Reach		
1	Method to identify target population	10 (100)
2	Inclusion criteria	10 (100)
3	Exclusion criteria	9 (90)
4	Participation rate	3 (30)
5	Representativeness	1 (10)
Average across reach indicators		6.6 (66)
Efficacy/effectiveness		
6	Measures or results for at least one follow-up	5 (50)
7	Intent-to-treat analysis	9 (90)
8	Quality-of-life or potential negative outcomes	6 (60)
9	Percent attrition	10 (100)
Average across efficacy/effectiveness indicators		7.5 (75)
Adoption		
10	Description of the intervention location	3 (30)
11	Description of staff who delivered the intervention	2 (20)
12	Method to identify staff who delivered the intervention (target delivery agent)	0 (0)
13	Level of expertise of the delivery agent	2 (20)
14	Inclusion and exclusion of the delivery agent or setting	0 (0)
15	Adoption rate of the delivery agent or setting	0 (0)
Average across adoption indicators		1.2 (12)
Implementation		
16	Intervention duration and frequency	10 (100)
17	Extent of the protocol delivered as intended	9 (90)
18	Measures of the cost of implementation	0 (0)
Average across implementation indicators		6.3 (63)
Maintenance		
19	Assessed outcomes at 6 months or following post intervention	5 (50)
20	Indicators of program-level maintenance	0 (0)
21	Measures of the cost of maintenance	0 (0)
Average across maintenance indicators		1.7 (17)

Discussion

Principal Findings

This review examined how digital features of internet-supported MBIs were related to the evidence of efficacy and intervention adherence and to which degree they informed capacity to translate into usual care using the RE-AIM framework. Since 2008, ten studies have examined the effects of an internet-delivered MBI on people with a chronic condition, with half of these studies published between 2014 and 2016. Findings indicated that internet-supported MBIs improved patient functioning for most outcome measures and were generally more effective than usual care or wait-list groups. Nevertheless, adherence to interventions was inconsistently and poorly defined and prevented robust comparison between studies. Self-guided interventions that allowed for flexible navigation of the program were as effective as facilitator-guided interventions, and more women with a chronic condition participated in an internet-supported MBI than men. This review also identified a number of reporting gaps within the RE-AIM framework, limiting the dissemination of internet-supported MBI research findings into practice.

Intervention Efficacy

Overall, internet-delivered MBIs were more effective than usual care or wait-list groups but not more effective than an active condition. This was demonstrated in three of the studies included in this review [44,47,48]. When an active control group was similar to the intervention delivery mode, time commitment, and attention, both the groups showed improved patient outcomes. This is an observation common to RCTs in general, where the type of control condition is known to affect study outcomes [50]. The extent to which participants in the active condition group were provided with a credible treatment rationale may have influenced their experience of that condition by generating positive expectations for improvement [51]. An intervention aiming to improve emotion-related outcomes using a mindfulness-based program is likely to trigger positive outcome expectations among individuals struggling to cope with a chronic condition. A usual care or wait-list group that serves as an untreated comparator would in those circumstances be more likely to experience a negative expectancy bias, which may translate into poorer outcomes [52]. Interestingly, a Web-based behavioral activation program was found to be more effective for people with severe depression than its MBI equivalent. The MBI was, however, more effective for individuals with lower levels of depression than the behavioral activation group [44]. As depressed individuals generally tend to experience concentration difficulties, distractibility, and problems with effortful cognitive processes [53], the authors of the latter study suggested that interventions requiring substantial cognitive functioning, such as attention control practice in MBIs, may not suit severely depressed participants. These findings suggest that the use of an active comparator could help to discern particular individual characteristics more sensitive to a mindfulness-based program.

Intervention Adherence

Adherence to the intervention was inconsistently defined across studies, which made comparison between studies difficult. This issue has previously been reported in other Web-based MBI reviews [6,7] and seems to endure across various types of behavioral Web-based research [54,55]. Adherence was broadly defined as the degree to which participants' behavior followed the recommendations from those delivering the program [56]. However, a single measure of adherence was not always appropriate for complex interventions, such as MBIs, which combined multiple modalities including educational components, meditation exercises, and mindfulness practice, to which participants may differentially adhere [55,56]. For example, in a study assessing the efficacy of a mindfulness manual delivered by email, adherence was defined as the extent to which the manual was read by participants [48]. The mindfulness program described in the manual contained an introduction to the concept of mindfulness and an explanation of how mindfulness can be practiced. In addition, a CD was provided for meditation exercises. Although nearly two-thirds of the participants reported having fully read the manual, it is unclear to what extent formal and informal mindfulness exercises were practiced. It is also unclear whether participants correctly understood the concept of mindfulness. However, it is also important to note that behavior change prompted by internet-based interventions may not require sustained or in-depth engagement with the program, as some users may require only a short period of intense engagement to initiate a habit or learn a skill, whereas others may need longer periods and a more personalized approach [57]. This may partly explain the incongruity between low-adherence rates and improved outcome measures observed in this review (eg, [43,46]). Participants who deviated from the recommended MBI structure may have still benefited from some aspects of the intervention. Given the important role of each modality in MBIs (ie, educational, informal practice, and meditation exercises), reporting on a multimodal measurement of adherence would provide an understanding of which aspects of MBI impact effectiveness.

Gender Disparity

Findings from this review were informed mainly by female participants (75%), which was slightly higher than those in in-person MBI studies where the average number of female participants was 71% [58]. Previous research showed that Web-based health-seeking behavior was reported to differ by gender, where women were more inclined to seek emotional and social support and affirmation of their health-related beliefs and men were interested mainly in health-related information [59]. Given the central focus of MBIs in health research is emotion regulation, these interventions may intuitively have a stronger appeal to women than to men. Gender differences in determinants and patterns of health behavior should be taken into account at the intervention design stage to accommodate male and female preferences.

Digital Characteristics

The digital characteristics listed in this review reflected those reported by the individual studies and were not exhaustive [8,9]. The majority of interventions were self-guided, delivered through a Web or mobile app, and allowed for flexible navigation of the program. Other features such as presentation strategies, including page design principles, average amount of text on pages, and the presence of hyperlinks to other resources may not only further our understanding of features influencing engagement and behavior change but also improve the reproducibility of the intervention in other contexts [60]. This is particularly relevant for interventions with low to no facilitator involvement, as the impact of the intervention relies primarily on digital features. Features such as the provision of the same information through various channels (text, audio, and video) to accommodate individual learning preferences [61], automated reminders to meditate, invitations to provide reflection on personal practice, automated progression feedback, and a range of meditation files to choose from could optimize intervention effects and inform learning preferences of different cohorts.

Internal and External Validity

This review used the RE-AIM framework to assess factors potentially hindering the translation of findings to clinical practice. Recognizing the RCT and quasi-experimental nature of the studies included, a focus on aspects related to internal validity was observed. Most indicators of reach, efficacy/effectiveness, and implementation were frequently

reported across studies. However, within these domains, essential indicators of generalizability, sustainability, and cost-effectiveness were rarely or never reported. For example, within the Reach domain, data related to the representativeness of study population were seldom reported, as most studies failed to address denominators such as populations from which settings, health professionals, and patients were drawn. The absence of this information hinders an analysis of the potential representation of the sample with the general population [13]. It is recommended that future studies report beyond the characteristics of study participants by comparing them with those of people declining to participate. If the recruitment process occurs in health care settings where patients are individually introduced to the study, then characteristics of people declining could be collected either directly from them by explaining the importance of this information or from the organization's database [13]. For online recruitment processes, existing databases such as population census data or national health surveys can be used to compare participants' demographic characteristics to people in the same community [13].

Most studies reported whether the trial focused on efficacy or effectiveness and, in general, reported on indicators pertaining to these domains. The distinction between these two approaches lay in the objective of the study. Efficacy (or explanatory) studies aim to investigate, under strictly controlled conditions, the difference between two treatments, whereas effectiveness (or pragmatic) studies investigate how an intervention fares in *real world* settings [62]. Despite this theoretical difference, in behavioral research, efficacy and effectiveness studies are generally conducted in *real world* settings such as university teaching hospitals or community health clinics, involving actual patients with real health problems being treated in real health care services [52]. The type of real-world setting needs to be described to allow adequate interpretation of study outcomes and inform generalizability. In this regard, intervention location selection, description, and adoption rates were rarely reported indicators. In addition, although half the studies in this review reported on follow-up outcome assessments at 6 months or more, possible program adaptation and maintenance requirements were never discussed. Reporting on factors influencing intervention adoption and maintenance will help inform resource allocation, potential for program dissemination, and replication of interventions in other settings. Furthermore, none of the studies reported on aspects related to cost other than for participatory incentives [42,46]. Dissemination plans need to be informed by cost incurred at both organizational and individual levels. Understanding cost incurred by recruitment (eg, staff qualifications needed to recruit participants), technology, and program adaptation and maintenance (eg, fixing technical problems) will help organizations adequately evaluate dissemination opportunities. Furthermore, knowing about program data usage and the type of service plans and digital devices best suited for the program will inform future cost to participants, which will have an impact on reach and effectiveness of the intervention [60]. Hence, future RCTs need to report resources needed to conduct the study, as insight into financial consequences will have practical implications for dissemination.

Of note was that less than half of the studies used a mindfulness measure. This is, however, similar to in-person MBI research where mindfulness outcomes were assessed in only 45% of the studies [63]. In this review, studies did not propose a clear potential mechanism of action for mindfulness [34]. The extent to which mindfulness influences intervention outcomes remains unclear. To understand the mechanistic role of MBIs, it is suggested that studies consistently use mindfulness measures and report related outcomes within a proposed theoretical framework [64].

Limitations

This review has several limitations that need to be considered when interpreting the results. First, data extraction was conducted by a single reviewer, which may have introduced some assessment bias. Second, digital characteristics described in this review were limited by the type of studies and populations included. Other study designs, such as pre-post studies, among, for example, a general population may describe additional features. Third, this review focused on the quality of reporting across the RE-AIM framework, which is different from the usual efficacy-based reviews that have a greater focus on the internal quality of the studies by performing risk of bias assessments [65]. However, two recent reviews of Web-based MBIs that focused on efficacy found that the quality of most studies was satisfactory and the interventions had a positive effect size on patient-reported outcomes [6,7]. These reviews both completed risk of bias assessments.

Conclusions

Findings from this review suggest that self-guided MBIs and those with minimal facilitator involvement can help alleviate the psychological burden associated with chronic disease. Future research is recommended to compare these types of interventions with other more established evidence-based therapies to identify the population groups that would benefit most from internet-supported MBIs.

Conflicts of Interest

None declared.

Multimedia Appendix 1

RE-AIM (reach, efficacy/effectiveness, adoption, implementation, maintenance) scoring instrument.

[PNG File, 54KB - mental_v5i3e53_app1.png]

References

1. Gotink RA, Chu P, Busschbach JJ, Benson H, Fricchione GL, Hunink MG. Standardised mindfulness-based interventions in healthcare: an overview of systematic reviews and meta-analyses of RCTs. PLoS One 2015 Apr;10(4):e0124344 [FREE Full text] [doi: 10.1371/journal.pone.0124344] [Medline: 25881019]
2. Chambers R, Gullone E, Allen NB. Mindful emotion regulation: an integrative review. Clin Psychol Rev 2009 Aug;29(6):560-572. [doi: 10.1016/j.cpr.2009.06.005] [Medline: 19632752]
3. Boggs JM, Beck A, Felder JN, Dimidjian S, Metcalf CA, Segal ZV. Web-based intervention in mindfulness meditation for reducing residual depressive symptoms and relapse prophylaxis: a qualitative study. J Med Internet Res 2014 Mar 24;16(3):e87 [FREE Full text] [doi: 10.2196/jmir.3129] [Medline: 24662625]
4. Waelde LC, Thompson JM, Robinson A, Iwanicki S. Trauma therapists' clinical applications, training, and personal practice of mindfulness and meditation. Mindfulness (N Y) 2016;7:622-629 [FREE Full text] [doi: 10.1007/s12671-016-0497-9] [Medline: 27217842]
5. Wahbeh H, Svalina MN, Oken BS. Group, one-on-one, or internet? Preferences for mindfulness meditation delivery format and their predictors. Open Med J 2014 Nov 28;1:66-74 [FREE Full text] [doi: 10.2174/1874220301401010066] [Medline: 27057260]
6. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
7. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303 [FREE Full text] [doi: 10.2196/jmir.7487] [Medline: 28860106]
8. Zhao D, Lustria ML, Hendrickse J. Systematic review of the information and communication technology features of web- and mobile-based psychoeducational interventions for depression. Patient Educ Couns 2017 Jun;100(6):1049-1072. [doi: 10.1016/j.pec.2017.01.004] [Medline: 28126383]
9. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
10. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003 Aug;93(8):1261-1267. [doi: 10.2105/AJPH.93.8.1261] [Medline: 12893608]
11. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
12. Gaglio B, Phillips SM, Heurtin-Roberts S, Sanchez MA, Glasgow RE. How pragmatic is it? Lessons learned using PRECIS and RE-AIM for determining pragmatic characteristics of research. Implement Sci 2014 Aug 28;9:96 [FREE Full text] [doi: 10.1186/s13012-014-0096-x] [Medline: 25163664]
13. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
16. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006 Jan;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
17. Linehan MM. Cognitive-Behavioral Treatment of Borderline Personality Disorder. New York: Guilford Publications, Inc; 1993.
18. Linehan MM. Skills Training Manual for Treating Borderline Personality Disorder. New York: Guilford Publications, Inc; 1993.
19. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008;26(2-4). [doi: 10.1080/15228830802094429]
20. Aihw. Chronic disease URL: https://www.aihw.gov.au/reports-statistics/health-conditions-disability-deaths/chronic-disease/about [accessed 2017-12-13] [WebCite Cache ID 6vfOEAC6w]
21. Helgeson VS, Zajdel M. Adjusting to chronic health conditions. Annu Rev Psychol 2017 Jan 3;68:545-571. [doi: 10.1146/annurev-psych-010416-044014] [Medline: 28051935]
22. 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 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

23. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012 Mar 9;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]
24. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from critical interpretive synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
25. Ugalde A, Haynes K, Boltong A, White V, Krishnasamy M, Schofield P, et al. Self-guided interventions for managing psychological distress in people with cancer - a systematic review. Patient Educ Couns 2017 May;100(5):846-857 [FREE Full text] [doi: 10.1016/j.pec.2016.12.009] [Medline: 28081937]
26. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]
27. Craike M, Hill B, Gaskin CJ, Skouteris H. Interventions to improve physical activity during pregnancy: a systematic review on issues of internal and external validity using the RE-AIM framework. BJOG 2017 Mar;124(4):573-583. [doi: 10.1111/1471-0528.14276] [Medline: 27571933]
28. Cuthbert CA, King-Shier K, Ruether D, Tapp DM, Culos-Reed SN. What is the state of the science on physical activity interventions for family caregivers? A systematic review and RE-AIM evaluation. J Phys Act Health 2017 Jul;14(7):578-595. [doi: 10.1123/jpah.2016-0280] [Medline: 28290737]
29. Matthews L, Kirk A, Macmillan F, Mutrie N. Can physical activity interventions for adults with type 2 diabetes be translated into practice settings? A systematic review using the RE-AIM framework. Transl Behav Med 2014 Mar;4(1):60-78 [FREE Full text] [doi: 10.1007/s13142-013-0235-y] [Medline: 24653777]
30. Eakin EG, Bull SS, Glasgow RE, Mason M. Reaching those most in need: a review of diabetes self-management interventions in disadvantaged populations. Diabetes Metab Res Rev 2002;18(1):26-35. [Medline: 11921415]
31. Allen K, Zoellner J, Motley M, Estabrooks PA. Understanding the internal and external validity of health literacy interventions: a systematic literature review using the RE-AIM framework. J Health Commun 2011 Sep;16 Suppl 3:55-72. [doi: 10.1080/10810730.2011.604381] [Medline: 21951243]
32. Glasgow RE, McKay HG, Piette JD, Reynolds KD. The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management? Patient Educ Couns 2001 Aug;44(2):119-127. [Medline: 11479052]
33. Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol 2014 Jan 2;5:e45 [FREE Full text] [doi: 10.1038/ctg.2013.13] [Medline: 24384867]
34. Edwards JR, Lambert LS. Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. Psychol Methods 2007 Mar;12(1):1-22. [doi: 10.1037/1082-989X.12.1.1] [Medline: 17402809]
35. Blackman KC, Zoellner J, Berrey LM, Alexander R, Fanning J, Hill JL, et al. Assessing the internal and external validity of mobile health physical activity promotion interventions: a systematic literature review using the RE-AIM framework. J Med Internet Res 2013 Oct 4;15(10):e224 [FREE Full text] [doi: 10.2196/jmir.2745] [Medline: 24095951]
36. Younge JO, Wery MF, Gotink RA, Utens EM, Michels M, Rizopoulos D, et al. Web-Based mindfulness intervention in heart disease: a randomized controlled trial. PLoS One 2015 Dec 7;10(12):e0143843 [FREE Full text] [doi: 10.1371/journal.pone.0143843] [Medline: 26641099]
37. Gotink RA, Younge JO, Wery MF, Utens EM, Michels M, Rizopoulos D, et al. Online mindfulness as a promising method to improve exercise capacity in heart disease: 12-month follow-up of a randomized controlled trial. PLoS One 2017 May 9;12(5):e0175923 [FREE Full text] [doi: 10.1371/journal.pone.0175923] [Medline: 28486559]
38. Zernicke KA, Campbell TS, Speca M, Ruff KM, Flowers S, Tamagawa R, et al. The eCALM trial: eTherapy for cancer applying mindfulness. Exploratory analyses of the associations between online mindfulness-based cancer recovery participation and changes in mood, stress symptoms, mindfulness, posttraumatic growth, and spirituality. Mindfulness (N Y) 2016 Oct;7(5):1071-1081. [doi: 10.1007/s12671-016-0545-5]
39. Thompson NJ, Patel AH, Selwa LM, Stoll SC, Begley CE, Johnson EK, et al. Expanding the efficacy of project UPLIFT: distance delivery of mindfulness-based depression prevention to people with epilepsy. J Consult Clin Psychol 2015 Apr;83(2):304-313 [FREE Full text] [doi: 10.1037/a0038404] [Medline: 25495361]
40. Thompson NJ, Walker ER, Obolensky N, Winning A, Barmon C, Diiorio C, et al. Distance delivery of mindfulness-based cognitive therapy for depression: project UPLIFT. Epilepsy Behav 2010 Nov;19(3):247-254. [doi: 10.1016/j.yebeh.2010.07.031] [Medline: 20851055]
41. Gardner-Nix J, Backman S, Barbati J, Grummitt J. Evaluating distance education of a mindfulness-based meditation programme for chronic pain management. J Telemed Telecare 2008;14(2):88-92. [doi: 10.1258/jtt.2007.070811] [Medline: 18348755]
42. Davis MC, Zautra AJ. An online mindfulness intervention targeting socioemotional regulation in fibromyalgia: results of a randomized controlled trial. Ann Behav Med 2013 Dec;46(3):273-284. [doi: 10.1007/s12160-013-9513-7] [Medline: 23670111]
43. Boettcher J, Aström V, Påhlsson D, Schenström O, Andersson G, Carlbring P. Internet-based mindfulness treatment for anxiety disorders: a randomized controlled trial. Behav Ther 2014 Mar;45(2):241-253 [FREE Full text] [doi: 10.1016/j.beth.2013.11.003] [Medline: 24491199]

44. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 9;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
45. Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. Psychosom Med 2014 May;76(4):257-267. [doi: 10.1097/PSY.0000000000000053] [Medline: 24804884]
46. Dimidjian S, Beck A, Felder JN, Boggs JM, Gallop R, Segal ZV. Web-based mindfulness-based cognitive therapy for reducing residual depressive symptoms: an open trial and quasi-experimental comparison to propensity score matched controls. Behav Res Ther 2014 Dec;63:83-89 [FREE Full text] [doi: 10.1016/j.brat.2014.09.004] [Medline: 25461782]
47. Dowd H, Hogan MJ, McGuire BE, Davis MC, Sarma KM, Fish RA, et al. Comparison of an online mindfulness-based cognitive therapy intervention with online pain management psychoeducation: a randomized controlled study. Clin J Pain 2015 Jun;31(6):517-527. [doi: 10.1097/AJP.0000000000000201] [Medline: 25565584]
48. Moritz S, Cludius B, Hottenrott B, Schneider BC, Saathoff K, Kuelz AK, et al. Mindfulness and relaxation treatment reduce depressive symptoms in individuals with psychosis. Eur Psychiatry 2015 Sep;30(6):709-714. [doi: 10.1016/j.eurpsy.2015.05.002] [Medline: 26163302]
49. Henriksson J, Wasara E, Rönnlund M. Effects of eight-week-web-based mindfulness training on pain intensity, pain acceptance, and life satisfaction in individuals with chronic pain. Psychol Rep 2016 Dec;119(3):586-607. [doi: 10.1177/0033294116675086] [Medline: 27780873]
50. Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychother Psychosom 2009;78(5):275-284. [doi: 10.1159/000228248] [Medline: 19602916]
51. Schwartz CE, Chesney MA, Irvine MJ, Keefe FJ. The control group dilemma in clinical research: applications for psychosocial and behavioral medicine trials. Psychosom Med 1997;59(4):362-371. [Medline: 9251155]
52. Freedland KE, Mohr DC, Davidson KW, Schwartz JE. Usual and unusual care: existing practice control groups in randomized controlled trials of behavioral interventions. Psychosom Med 2011 May;73(4):323-335 [FREE Full text] [doi: 10.1097/PSY.0b013e318218e1fb] [Medline: 21536837]
53. Gotlib IH, Joormann J. Cognition and depression: current status and future directions. Annu Rev Clin Psychol 2010 Mar;6:285-312 [FREE Full text] [doi: 10.1146/annurev.clinpsy.121208.131305] [Medline: 20192795]
54. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
55. Ryan C, Bergin M, Wells JS. Theoretical perspectives of adherence to web-based interventions: a scoping review. Int J Behav Med 2018 Feb;25(1):17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
56. Vitolins MZ, Rand CS, Rapp SR, Ribisl PM, Sevick MA. Measuring adherence to behavioral and medical interventions. Control Clin Trials 2000 Oct;21(5 Suppl):188S-194S. [Medline: 11018574]
57. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
58. Bodenlos JS, Strang K, Gray-Bauer R, Faherty A, Ashdown BK. Male representation in randomized clinical trials of mindfulness-based therapies. Mindfulness 2017;8(2):259-265. [doi: 10.1007/s12671-016-0646-1]
59. Baumann E, Czerwinski F, Reifegerste D. Gender-specific determinants and patterns of online health information seeking: results from a representative German health survey. J Med Internet Res 2017 Apr 4;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
60. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
61. Russell SS. An overview of adult-learning processes. Urol Nurs 2006 Oct;26(5):349-52, 370. [Medline: 17078322]
62. Kowalski CJ. Pragmatic problems with clinical equipoise. Perspect Biol Med 2010;53(2):161-173. [doi: 10.1353/pbm.0.0161] [Medline: 20495255]
63. Khoury B, Lecomte T, Fortin G, Masse M, Therien P, Bouchard V, et al. Mindfulness-based therapy: a comprehensive meta-analysis. Clin Psychol Rev 2013 Aug;33(6):763-771. [doi: 10.1016/j.cpr.2013.05.005] [Medline: 23796855]
64. Davidson RJ, Kaszniak AW. Conceptual and methodological issues in research on mindfulness and meditation. Am Psychol 2015 Oct;70(7):581-592 [FREE Full text] [doi: 10.1037/a0039512] [Medline: 26436310]
65. Charrois TL. Systematic reviews: what do you need to know to get started? Can J Hosp Pharm 2015;68(2):144-148 [FREE Full text] [Medline: 25964686]

Abbreviations

C: control group

FU: follow-up
I: intervention
MBI: mindfulness-based intervention
MBCPM: mindfulness-based chronic pain management
MBCT: mindfulness-based cognitive therapy
MBCR: mindfulness-based cancer recovery
MBSR: mindfulness-based stress reduction
MSER: mindful socioemotional regulation
NR: not reported
PICOS: intervention, comparison, control, or comparator, outcome, and study type
QE: quasi-experimental
RCT: randomized controlled trial
RE-AIM: reach, efficacy/effectiveness, adoption, implementation, maintenance

Edited by G Eysenbach; submitted 14.12.17; peer-reviewed by K Toivonen, J Cockle-Hearne, YF Tsai, D Reifegerste; comments to author 18.03.18; revised version received 07.05.18; accepted 21.06.18; published 21.08.18.

Please cite as:
Russell L, Ugalde A, Milne D, Austin D, Livingston PM
Digital Characteristics and Dissemination Indicators to Optimize Delivery of Internet-Supported Mindfulness-Based Interventions for People With a Chronic Condition: Systematic Review
JMIR Ment Health 2018;5(3):e53
URL: http://mental.jmir.org/2018/3/e53/
doi:10.2196/mental.9645
PMID:30131317

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org