JMIR Mental Health

Contents

Original Paper

Experiences of a Web-Based Quality of Life Self-Monitoring Tool for Individuals With Bipolar Disorder: A Qualitative Exploration

Emma Morton[1], BSc, PhD; Rachelle Hole[2], BA, BSW, MSW, PhD; Greg Murray[3], BA, BSc, MPsych, PhD; Simone Buzwell[3], BA, PhD; Erin Michalak[1], BSc, PhD

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

[2]School of Social Work, University of British Columbia, Okanagan, BC, Canada

[3]Department of Psychological Sciences, Swinburne University of Technology, Hawthorn, Australia

Corresponding Author:
Erin Michalak, BSc, PhD
Department of Psychiatry
University of British Columbia
420-5950 University Blvd
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 604 827 3393
Email: erin.michalak@ubc.ca

Abstract

Background: Self-monitoring of symptoms is a cornerstone of psychological interventions in bipolar disorder (BD), but individuals with lived experience also value tracking holistic outcomes, such as quality of life (QoL). Importantly, self-monitoring is not always experienced positively by people with BD and may have lower than expected rates of engagement. Therefore, before progressing into QoL tracking tools, it is important to explore user perspectives to identify possible risks and benefits, optimal methods to support engagement, and possible avenues to integrate QoL self-monitoring practices into clinical work.

Objective: This study aimed to conduct a qualitative exploration of how individuals with BD engaged with a Web-based version of a BD-specific QoL self-monitoring instrument, the QoL tool.

Methods: A total of 43 individuals with BD engaged with a self-management intervention with an optional Web-based QoL self-assessment tool as part of an overarching mixed method study. Individuals were later interviewed about personal experiences of engagement with the intervention, including experiences of gauging their own QoL. A thematic analysis was used to identify salient aspects of the experience of QoL self-monitoring in BD.

Results: In total, 4 categories describing people's experiences of QoL self-monitoring were identified: (1) breadth of QoL monitoring, (2) highlighting the positive, (3) connecting self-monitoring to action, and (4) self-directed patterns of use.

Conclusions: The findings of this research generate novel insights into ways in which individuals with BD experience the Web-based QoL self-assessment tool. The value of tracking the breadth of domains was an overarching aspect, facilitating the identification of both areas of strength and life domains in need of intervention. Importantly, monitoring QoL appeared to have an inherently therapeutic quality, through validating flourishing areas and reinforcing self-management efforts. This contrasts the evidence suggesting that symptom tracking may be distressing because of its focus on negative experiences and positions QoL as a valuable adjunctive target of observation in BD. Flexibility and personalization of use of the QoL tool were key to engagement, informing considerations for health care providers wishing to support self-monitoring and future research into Web- or mobile phone–based apps.

(JMIR Ment Health 2019;6(12):e16121)   doi:10.2196/16121

KEYWORDS

Introduction

Background

Bipolar disorder (BD) is a chronic mood disorder characterized by periods of mania (irritable or elevated mood with increased activity) and depression, experienced by up to 2.4% of the global population [1]. Self-monitoring of symptoms is an active ingredient of psychosocial therapies for BD [2], and it is recommended as an adjunctive strategy in treatment guidelines [3]. People with lived experience of BD have expressed interest in tracking additional life domains. Quality of life (QoL), a holistic outcome, has strong potential to meet this market need. Before progressing into QoL tracking apps, it is important to attend to the evidence suggesting that self-monitoring in BD is not always experienced positively. Attention to symptoms may be a distressing reminder of being unwell [4] and may reinforce depressive symptoms [5]. Given its broad focus and attention to both positive and negative experiences, QoL may be an acceptable target for self-monitoring and play a valuable complementary role to symptom tracking. This qualitative study is an initial step toward understanding self-monitoring of QoL in people with BD by exploring user experiences of a Web-based BD-specific QoL measurement instrument.

Self-monitoring of symptoms is a cornerstone of psychological therapies for BD used to support self-management, detect early warning signs of relapse, and guide appropriate intervention. Traditionally, such interventions have focused on sleep and mood, which have been shown to improve time to recurrence of mood episodes, rates of hospitalization, and functioning [6]. However, there is evidence to suggest that people with BD are interested in tracking other life domains. A survey of self-monitoring strategies used by people with BD indicated that although mood and sleep were most commonly monitored, individuals also tracked a variety of other areas including finances, social interactions, substance use, household management, pet care, and leisure time [7]. Individuals created *elaborate* methods to track multiple indicators, suggesting a market need for apps facilitating simultaneous tracking of diverse life domains. QoL is a broad construct taken to represent aspects of functioning and satisfaction in occupational, environmental, social, physical, and psychological aspects of life [8]. QoL may, therefore, represent a prime outcome that addresses the desire for more inclusive forms of self-monitoring in BD. Furthermore, individuals with BD have nominated QoL as an outcome that is prioritized alongside or even *above* symptom reduction [9-14]. Increasing focus has been placed on QoL-focused self-management strategies and psychosocial interventions for BD [15,16], of which, QoL tracking would form a natural complement. Although studies on the validity, feasibility, and acceptability of routine QoL self-assessment have been conducted in oncology [17-19], this experience remains poorly understood in mental health.

Despite research indicating positive effects of self-monitoring in BD, those with lived experience may be ambivalent or inconsistent about daily assessment. Rates of adherence to self-monitoring interventions in BD vary widely, with missing data rates between 6% and 58% for electronic methods [20,21].

Although some disengagement is expected given the effort required [22], this may not fully explain adherence rates. Studies in BD and other psychiatric conditions suggest that there are additional psychological costs to self-monitoring when living with a chronic mental illness. One such qualitative study specifically explored experiences of using the Life Chart Methodology (LCM), a daily record of manic and depressed mood, sleep duration, number of mood changes, and medication use [23], in individuals with bipolar I disorder (BD-I) [4]. The usage of the LCM was perceived by almost half (43%) of the participants as an unpleasant reminder that they were living with BD, forcing confrontation with the limitations of illness. Some believed that focusing on negative experiences increased their depression, and this was associated with a reluctance to use the LCM. The authors theorized that aversive experiences of the LCM were linked the psychological threat of losing the identity of the *healthy self*. Although quantitative research has not explored these experiences on a larger scale, there is some support for the idea that self-monitoring may reinforce negative experiences. Indeed, a study of daily BD mood tracking using mobile phone technology was shown to increase depressive symptoms in the intervention group [5]. In addition, in digital health interventions for psychosis, qualitative studies and adverse event reports suggest that routine self-assessment of symptoms may increase rumination on negative experiences, generate fear of relapse when negative changes in mental health are observed, or trigger distressing memories of episodes of psychosis [24,25].

Objective

It is not known whether QoL self-monitoring would be associated with a similar impact as symptom tracking in BD. Qualitative exploration of such applications may answer questions that quantitative studies of feasibility and acceptability may not, including possible risks and benefits, optimal integration of QoL data into clinical practice, and strategies to support engagement with self-assessment tools [25]. The aim of this study was to conduct a qualitative exploration into how individuals with BD engaged with a Web-based version of a BD-specific QoL self-monitoring instrument, the QoL tool [26].
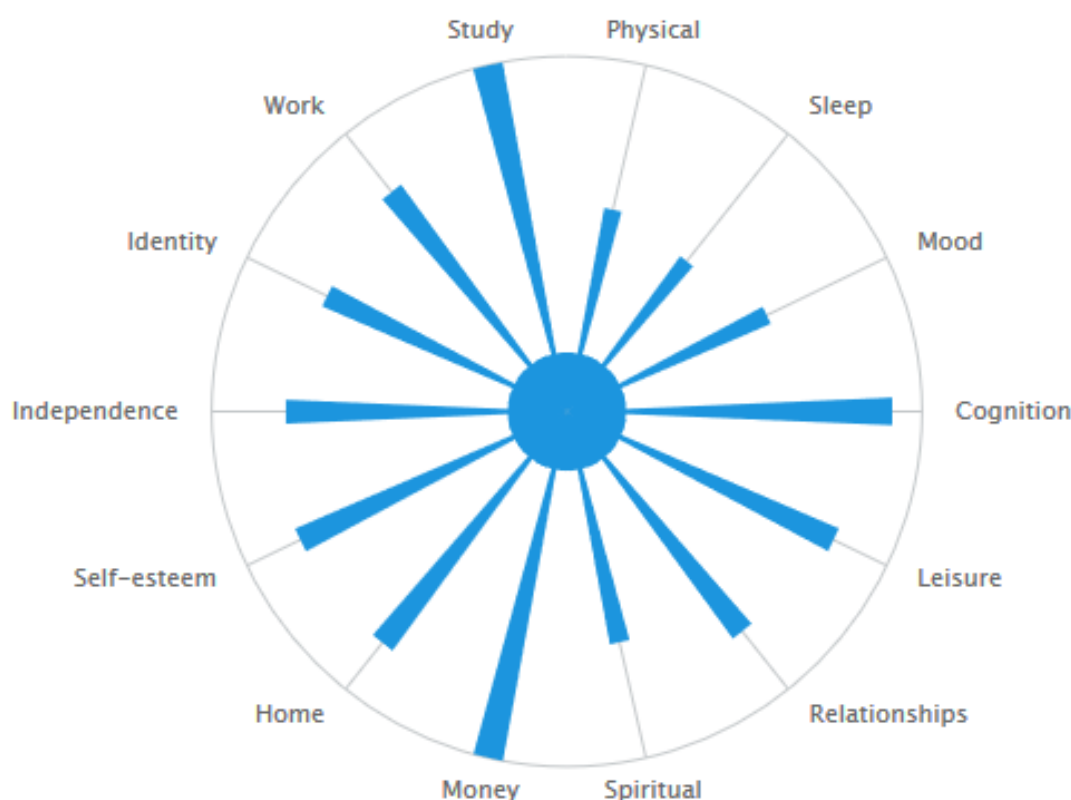
Methods

Design

By using descriptive qualitative methods [27], this study was conducted within the context of a broader mixed method (ie, both quantitative and qualitative) project investigating the impact of self-management in BD [16]. The aim of the overarching project was to investigate the effectiveness of a website (Bipolar Wellness Centre [28]) and various associated information delivery modalities for communicating and encouraging the use of various self-management strategies (for a detailed description of these, see Michalak et al. [16]). The Bipolar Wellness Centre linked to the externally hosted QoL tool [29], which contains a Web-based version of a BD-specific QoL self-report measure developed through a Community-Based Participatory Research process. Life domains important to the experience of QoL in BD were elicited through interviews with individuals with BD, their family members, and health care providers [30]. The final

instrument assesses satisfaction in the last week across domains of relationships, home duties, leisure activities, physical health, sleep, mood, money, independence, spirituality, self-esteem, cognition, identity, and work or study, if relevant [26]. Questionnaire items are positively framed, that is, respondents are asked about their level of enjoyment and positive experiences across domains, rather than distress or problems. Results from the QoL tool are represented graphically, and the quantitative summary of scores gives the user an optional link to the corresponding page of the Bipolar Wellness Centre (Figure 1).

An optional introductory video describes these features and the developer's recommendations for use (completing the measure weekly). Participants in the primary mixed method study were free to choose whether they would complete the QoL tool. To address the aim of this study to explore experiences of self-monitoring QoL, the focus was on the participants' description of engagement with the QoL tool rather than specific self-management strategies or aspects of the Bipolar Wellness Centre that they may have also used.

Figure 1. Representation of quality of life tool results.

Recruitment

The participants for this study were a subsample of those in an overarching project who agreed to participate in a follow-up qualitative telephone interview. Participants in the primary project were required to meet the following criteria: (1) aged 19 years or above, (2) able to communicate in English, (3) able to provide informed consent, (4) a resident of Canada, and (5) have a self-reported diagnosis of BD. Self-reported diagnoses of BD was considered sufficient, given that this is typically congruent with the results of formal diagnostic interviews [31].

Recruitment in the overarching project was conducted via the following: (1) notices were sent to participants in prior BD studies in the Mood Disorders Centre, Department of Psychiatry (University of British Columbia), who consented to be contacted in the future regarding new studies of potential interest, (2) health care providers affiliated with the Collaborative Research Team for Psychosocial Issues in BDs (CREST.BD) network

were provided with information about this project and asked to place informational leaflets in their waiting room, and (3) advertisements in print (eg, community newspapers) and Web-based (eg, blogs) media.

At enrollment in the project, individuals indicated their consent to be contacted at a later date regarding participation in a telephone interview. For the qualitative arm, purposive sampling was used to identify consenting participants, who were then contacted via email and invited to participate in the qualitative interview. Participants received a Can $20 honorarium in the form of a gift certificate as compensation for time spent in the interview.

Purposeful criterion sampling [32] was used to establish a sample that reflected the diversity of the primary project sample, namely, demographics and engagement with the various components of the Bipolar Wellness Centre. Participants were specifically recruited to ensure representation of various age

groups, gender, BD subtype, and experiences with the spectrum of self-management strategies across the information delivery modalities. For the purposes of this study, these sampling criteria were not applied in the analytic framework or the description of the findings, as focus was on experiences of engaging with the QoL tool.

Qualitative Interview

A semistructured interview schedule was developed based on 3 sets of topics informed by the overarching project aim (engagement with information on self-management strategies, implementation of self-management strategies, and perceived impact on QoL). Main findings have been published elsewhere [16,33,34]. Germane to this subanalysis, participants were asked if they had used the QoL tool or had any plans to use it in future. They were also asked whether using the QoL tool had any impact on their engagement with the Bipolar Wellness Centre (eg, "Did completing the QoL tool have any impact on the kinds of changes you thought about making in your life?") or their perceived QoL (eg, "Did completing the QoL tool impact your QoL in any way?"). Probes and reflective listening were used to elicit depth in participant responses.

Procedure

The Behavior Research Ethics Board of the University of British Columbia approved the study. All participants received written information on the study and gave written consent to be contacted for an interview. Data in the study were treated confidentially, and transcripts were deidentified.

Consenting participants were contacted via email approximately 2 weeks after participating in the self-management intervention to schedule a telephone interview. In total, 67 participants were invited to participate in the qualitative interview; of these, 24 declined to participate or did not respond. A total of 43 interviews were conducted by the first author. Interviews lasted from 20 to 70 min (mean 39.4 min, SD 11.2). They were digitally recorded and transcribed verbatim by the first author (24/43, 55%) or research assistants (checked by the first author for accuracy; 19/43, 44%).

Data Analysis

Braun and Clark's [35] guidelines for thematic analysis were followed. A thematic analysis provides a flexible research tool that can provide a rich, detailed, and complex account of patterns across an entire dataset. As the primary aim of the interview schedule was to explore experiences of the self-management intervention and perceptions of QoL, there was relatively limited discussion of participants' experiences of engaging with the QoL tool. Consequently, findings are presented at the level of *descriptive, literal categories* rather than interpretative and abstracted themes [35].

The first author was familiarized with the data through the process of transcribing interviews and re-reading transcripts. Data were assigned brief descriptive codes (meaning units) in the qualitative data management software NVivo (QSR International, 2016). Codes were examined, and descriptive categories were generated [35]. The content of categories was reviewed for coherency, and transcripts were revisited as

categories were developed to ensure that all relevant data were adequately captured and described. The essence of the most salient categories in relation to this aim has been summarized with illustrative transcript extracts. To reduce the risk of bias and address analytic validity, coauthor EM reviewed both the descriptive accounts of categories and transcripts for coherency, with disagreements resolved via consensus.

Results

Participants

A total of 43 participants were interviewed for the qualitative study (42.2% of the primary study sample). The modal age range was 45 to 54 years. Most participants (30/43, 69%) were female. The most frequently reported diagnosis was BD-I (24/43, 55%), followed by bipolar II disorder (BD-II) (16/43, 37%). A participant reported a diagnosis of rapid cycling BD-II, and 2 participants reported a diagnosis of bipolar disorder not otherwise specified (BD-NOS).

Overview of Key Findings

Of the interviewed sample, 74% (n=32/43) reported using the QoL tool. In total, 4 key categories are described: (1) breadth of QoL monitoring, (2) highlighting the positive, (3) connecting self-monitoring to action, and (4) self-directed patterns of use.

Breadth of Quality of Life Monitoring

A predominant experience of QoL self-monitoring was the value of reporting satisfaction across a range of life domains. The QoL tool was described as supporting users to attend to a greater range of life domains than they may otherwise monitor, by reminding people of important areas and increasing the granularity with which they appraised their circumstances:

> It has helped me look at the different domains and pay attention... it's all one big thing, but now that it's in domain names it's kind of easier to organize and pay attention to those domains. [Male, aged 35-44 years, BD-I]

Receiving a breakdown of scores across domains was described as helping people consider the influence of each aspect of QoL separately, which could facilitate the identification of areas in need of improvement that are otherwise masked by an overall feeling of well-being:

> It's really good having that separation into categories in terms of thinking of each one independently of the other... even when you feel like overall you're doing well, there could be one thing that, when you actually sit down and honestly think about it, you may discover that's an area you may need to work on more. [Male, aged 45-54 years, BD-II]

For some participants, the ability to parse the influence of various aspects of QoL supported them in identifying the causes of mood changes. A participant described using QoL tool data to facilitate discussions with their psychiatrist about possible relapse prodromes:

> If something is starting to get wonky... it's easier when you have to talk to your psychiatrist because you can

bring it in and say "this is what's going on, this area is getting affected, these areas are getting affected." [Female, aged 35-44 years, BD-I]

In some cases, the range of domains covered by the QoL tool was explicitly contrasted to traditional forms of self-monitoring in BD. A participant noted that it was difficult to provide a single rating on a mood diary that captured the variety of their day-to-day experiences:

I always hated the mood diaries, it seems too constricting to just rate my mood on a scale of 1 to 10 for one day. [Female, aged 25-34 years, BD-I]

The breakdown of results of the QoL tool also reduced the burden of interpreting changes:

The daily mood chart, you've got to interpret it, you've got to sit there and look back over it... and this tool it's right there. [Male, aged 55-64 years, BD-I]

The breadth of domains was not universally appreciated; in an instance, monitoring traditional outcomes was preferred. A participant described intentions to track their sole area of concern (sleep) instead of using the QoL tool.

Highlighting the Positive

A number of participants reported that the QoL tool served an important function in highlighting life domains where they were flourishing. Individuals described the QoL tool results as drawing their attention to areas of strength, which was often accompanied by a sense of appreciation or a positive affirmation that one was on the right track:

I feel like the tool is good even when you're doing well because, any kind of positive, anything that makes me feel good about myself, is always positive... That's kind of what it feels like, that encouragement, that—yes, you're doing good, you're doing a good job. [Male, aged 45-54 years, BD-II]

Related to the importance of monitoring a breadth of QoL domains, individuals appreciated being able to note strengths at the same time as identifying areas in need of additional support:

Here's where I'm satisfied, here's where I'm not satisfied... there are some things I want to change, but I wasn't mad at myself when I looked at the results. [Female, aged 25-34 years, BD-II]

Tracking strengths could also play a powerful role in identity. An individual described a sense of reclaiming who they were beyond the *sick role* of bipolar by noticing the range of areas where they were performing well:

It made me feel really good actually, because I knew I was on the right track... yes, I suffer from bipolar, and now it's [participant's name] and I'm going to work, I'm functioning, full time hours, and able to handle a lot more than I used to. [Female, aged 35-44 years, BD-I]

In some cases, reviewing results prompted a positive reappraisal of circumstances. For example, a participant described reflecting

on their relationships after they received a higher domain score than predicted:

I saw that and thought... maybe it's not such a horrible situation, and I think it's better than I might have thought. [Female, aged 45-54 years, BD-I]

Noting positive results was also described as drawing attention to and encouraging continued application of self-management strategies that were supporting the areas of good QoL. This is exemplified by a participant who added additional physical activities to their schedule after reflecting on the positive impact of their existing routine:

It made me appreciate what I was already doing and I kept doing it... but then it made me think about how that is helping me, so I started swimming as well. [Female, aged 25-34 years, BD-I]

The connection between QoL tracking and self-management practices is explored in detail in the following section.

Connecting Self-Monitoring to Action

Over half of those who used the QoL tool described it as guiding their self-management efforts (n=17). Individuals found QoL self-monitoring enabled them to identify areas where they scored lower, which prompted them to either implement self-management strategies or focus their efforts researching ways to improve that specific domain:

Doing the Quality of Life tool is what made me think about my social activities being very weak... based on that awareness I chose to take action on it. [Female, aged 45-54 years, BD-II]

Aspects of the display of QoL tool results were flagged as important in connecting self-monitoring to action. For some individuals, the visual aspect of the QoL tool *results* display was particularly important in identifying areas in need of attention. In such cases, the idea of an *unbalanced* wheel was described, as in the following account:

It also made me realize how unbalanced I could be... I was okay in many aspects, and going back to relationships for me, was really bad on that aspect. I kind of knew it but when I saw it represented that way with the spokes, I felt I had to work on this. [Male, aged 45-54 years, BD-I]

A participant described the numerical breakdown of scores as an objective indicator motivating them to take action:

When I would get a score of 11 out of 20 on something, it made me read more, investigate further on that particular area because it pointed out to me in a very quantitative way, that it's not just that you sort of think there's a problem or something you want to improve on, that it actually shows you with numbers, and that's pretty concrete. [Female, aged 65-74 years, BD-II]

Another aspect specific of the QoL tool (vs the paper-and-pencil version of the instrument) resides in its links to the Bipolar Wellness Centre. A small number of people explicitly described using the links in their results summary (see Figure 1) to access

XSL•FO

RenderX

information about relevant self-management strategies for domains of interest:

> *Afterwards, there's an article on the website itself that talks about the areas and links to all these other resources... that really helped.* [Female, aged 25-34 years, BD-I]

For some participants, using the QoL tool on an ongoing basis itself became a self-management strategy because of its ability to remind them to take action:

> *I do need to be reminded of things every so often, so if I took that quiz more often it would probably help me remind myself.* [Female, aged 25-34 years, BD-I]

Ongoing plans for QoL self-monitoring are discussed in more detail in the following section.

Self-Directed Patterns of Use

When asked about their use of the QoL tool (in the weeks since the start of the intervention and plans for future use), most people talked about tracking QoL in an ongoing sense, but the intended patterns of use varied. Some participants described plans for routine use on a weekly, biweekly, or monthly basis. In such cases, QoL self-monitoring was sometimes described as a preventative strategy, for example "to track things before they get so serious" (female, aged 45-54 years, BD-I). Others spoke about the importance of obtaining enough data to provide insight into patterns:

> *I think it's important to do the Quality of Life tool on a fairly regular basis, because that's the only way you can identify changes or trends, not just when you're at a peak or valley.* [Female, aged 45-54 years, BD-II]

For others, regular use was important to keep self-management efforts on track:

> *I try to do it weekly, and it's a reminder that I have to look after all aspects.* [Male, aged 55-64 years, BD-I]

Others planned on using the QoL tool infrequently, such as every few months. This was sometimes described as *checking in* on an ad hoc basis to see if progress could be detected:

> *I certainly will do it again, maybe in a few months again to see if anything has changed, has improved.* [Male, aged 35-44 years, BD-II]

For others, QoL tracking could be deployed in response to mood changes to gather data about the cause or identify helpful strategies:

> *There's a tool that acts as a checklist for me: when I don't feel great, I can go over that and it helps me.* [Male, aged 45-54 years, BD-I]

Finally, QoL self-monitoring may only occur if usual forms of gauging wellness were not available (eg, the opinion of a trusted health care provider when going on holiday).

Participants identified ways in which the QoL tool could be altered to support their ongoing use, such as utilizing prompts to self-monitor if the website had not been accessed in some

time or providing a quantification of change since the last completed questionnaire. Interestingly, 2 people took matters into their own hands and adapted the questionnaire to suit their preferences: 1 created a laminated version of the display of the QoL tool results (see Figure 1) on which they marked their daily self-assessment. Another participant created a Microsoft Excel table aggregating their scores from the QoL tool, along with a *diary* column to help pinpoint the cause of changes in QoL. Personalization appeared to support engagement with QoL self-monitoring; in both cases, individuals described the QoL tool as a key component of their ongoing self-management plan:

> *It's not just a bunch of scores. This is at the heart of my entire self-management strategy.* [Male, aged 35-44 years, BD-I]

Discussion

Principal Findings

This qualitative study explores the subjective experience of QoL self-monitoring in BD by interviewing individuals who had utilized a BD-specific QoL measure, the QoL tool. This investigation adds to the growing literature on digital monitoring in BD but is the first to comment on a QoL-specific app. Furthermore, although this qualitative study was conducted with a sample of individuals who self-selected into using the QoL tool in the context of a self-management intervention, it has implications for supporting self-reflection and utilizing subjective QoL data in clinical practice.

The breadth of life domains assessed by the QoL tool was identified as an important aspect of engagement with self-monitoring and appears fundamental to other key experiences described herein: assessing a range of domains increases the chances of individuals identifying the areas of strength alongside those in need of remedial action. The perceived value of tracking a range of domains accords with the findings that people with BD are already attempting to log multiple indicators beyond symptoms [7]; however, few apps exist to support this. A systematic review of apps for BD showed that the majority of self-monitoring apps track symptoms, sleep, and medication use [36]. Owing to interest in gauging holistic outcomes and the lack of technologies to support this, the positive responses to a Web-based QoL self-report measure documented in this study validate further study and development of QoL self-assessment tools. Furthermore, given increasing attention to mobile phone–based monitoring of symptoms [5,37,38], complementary QoL-focused apps may be a useful focus of further development.

A key aspect of the experience of QoL self-monitoring was the positive nature of using results to identify the areas of strength. This may be expected given that the QoL tool includes positive rather than negative indicators (ie, individuals are asked about enjoyment and satisfaction, rather than distress and dysfunction). It has been argued that scales, which include positive indicators, enable the detection of flourishing areas, whereas those composed solely of negative indicators tend to underestimate a person's QoL [39]. Although this does not preclude respondents attending to negative aspects of their circumstances, it may be argued that a broad, positively framed questionnaire

is more likely to detect the areas of strength for 2 reasons. First, in accordance with recovery perspectives [40], the broader lens of QoL places de facto emphasis on life domains where people may experience satisfaction and meaning despite the limitations of illness. Second, the QoL tool has positively framed item wording that may promote attention to strengths. Positively framed questions have been used therapeutically to encourage respondents to adopt a shared perspective with the questioner; if a question asks about strengths, individuals will seek evidence of this from their own lives [41]. There are important clinical implications affirming the experience of QoL self-monitoring in this study. First, QoL-focused monitoring may address a noted limitation of symptom tracking, which is the reinforcement of negative symptoms. An analysis of app store reviews of cognitive behavioral therapy apps for depression revealed the primary criticism was the sole focus on negative experiences, with users worrying that this could reinforce negative thinking patterns [42]. Similarly, qualitative studies in BD and psychosis highlights user concerns that attention to symptoms reinforces negative affect [4,25]. Conversely, as suggested by the experiences of participants in this study, tracking *positive* experiences may amplify their emotional and psychological impact or induce a positive reframing of circumstances. This is supported by quantitative research showing an experience sampling intervention focusing on positive affect that reduced depressive symptoms in a sample of patients with major depressive disorder [43,44]. Future quantitative comparisons of the psychological and emotive impact of mood charting versus QoL self-monitoring may help further illuminate their optimal therapeutic apps, such as identifying individuals who may be more likely to be negatively impacted by mood charting or the adjunctive benefits of the more positively oriented QoL monitoring. The second potential clinical app of drawing attention to strengths in broader life domains via reflecting on QoL is fostering the development of a holistic sense of self, a process that is thought to be disrupted in BD [45,46]. Although traditional outcome tracking may also assist with this by helping understand the shifts in mood [47], some individuals may find that it threatens other valued aspects of identity, such as the concept of the *healthy self* [4]. At least one individual in this study expressed a sense that drawing attention to areas of strength through QoL self-monitoring assisted with recovering their identity beyond *someone with bipolar*, which accords with the suggestion that developing balanced and less pessimistic views of the self is a key therapeutic mechanism of psychosocial interventions in BD [2]. Taken together, this suggests possible inherently therapeutic applications of QoL self-monitoring (that is, beyond its possible role in guiding self-management efforts, discussed below), which future studies could explore in more detail.

The potential role of QoL self-monitoring in guiding and encouraging self-management is an important area of future investigation, particularly, as self-management is an increasingly recognized component of treatment in BD [3]. Individuals in this study reported using QoL tool data to identify priority areas for implementing self-directed wellness strategies. This application could be extended to the support of self-management practices and the identification of treatment goals in the clinical context, which has been suggested to improve therapeutic

alliance [48] and motivation to engage in treatment [49]. However, this study was generated in the context of self-directed QoL self-monitoring; experiences of sharing data with treating teams were not explored in detail nor were investigations conducted into how health care providers made use of these data. It has been noted that the health care provider's uncertainty about how to interpret and respond to self-reported data may pose a risk to therapeutic alliance [7] or negatively impact engagement in self-monitoring [25]. Therefore, complementary research on optimal methods for integrating patient-reported QoL data into treatment planning and outcome assessment in the clinical context is needed.

The optimal frequency of self-monitoring in BD is unknown [50]. This study examined self-directed patterns of use, highlighting 2 groups: individuals who engaged with the QoL tool routinely and those who planned to *check in* on an ad hoc basis. Disengagement is common with digital health interventions [22], and although engagement was not formally assessed over an extended period in this study, reported plans for use are promising. Potentially, the lack of a required frequency of QoL monitoring and the strength-focused nature encourage ongoing use, which warrants further investigation. Of note, some participants expressed desires for additional features or themselves made modifications to the QoL tool, suggesting the following: (1) a need for lived experience involvement in the development of future iterations of digital QoL tracking apps (such as user-centered design [51]) and (2) health care providers initiating conversations with users about how self-monitoring can be supported or adapted to meet their unique needs and preferences.

Limitations

First, the opportunistic sample was drawn from an evaluation of self-management knowledge translation, and as such, they may have had more positive views of QoL self-monitoring and self-management. It is not known whether the experiences described here would generalize to individuals not engaged in psychosocial interventions. In addition, as the overarching study aimed to specifically evaluate the Bipolar Wellness Centre, quantitative data on the frequency of use of the QoL tool were not available to objectively verify self-reported patterns of use. Future research should be conducted to quantify the rates of use of the QoL tool, evaluate the potential for QoL tracking to support engagement in self-management, and investigate ways to support engagement in self-monitoring in a broader sample. Furthermore, purposive sampling of individuals who disengage from psychosocial interventions or QoL self-monitoring may illuminate negative experiences in more detail than this study. Second, the use of the QoL tool was not the primary focus of the interviews, and the data were less rich than the themes explored regarding engagement with the self-management intervention [34] and perceptions of own QoL [33]. We did not probe in depth for prior experiences of mood monitoring or ask participants to contrast this to their experiences of QoL tracking. To address the risk of overinterpretation of the evidence, analysis of the qualitative data was limited to descriptive, literal categories, rather than risk bias by generating interpretative themes [52].

Conclusions

For the first time, individuals with BD have commented on the experience of engaging with a Web-based tool for QoL self-monitoring. This study highlights the unique characteristics of QoL tracking that were valued by the participants, including the breadth of domains assessed, the ability to affirm strengths, and the link between reflecting on QoL tool results and actions to improve one's circumstances. A variety of patterns of use were described, with an emphasis on flexibility. It is concluded that QoL self-monitoring shows great potential to support recovery-oriented and self-management interventions and may have an inherently therapeutic quality relevant to the treatment of BD.

Acknowledgments

Authors' Contributions

EM contributed to the design of the study, conducted the data collection and analysis, and wrote the paper. EEM, RH, and GM contributed to the design of the study, data analysis, and revisions of the paper. SB contributed to the data analysis and revisions of the paper.

Conflicts of Interest

None declared.

References

1. Merikangas KR, Jin R, He J, Kessler RC, Lee S, Sampson NA, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. Arch Gen Psychiatry 2011 Mar;68(3):241-251 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2011.12] [Medline: 21383262]

2. Miklowitz DJ, Scott J. Psychosocial treatments for bipolar disorder: cost-effectiveness, mediating mechanisms, and future directions. Bipolar Disord 2009 Jun;11(Suppl 2):110-122. [doi: 10.1111/j.1399-5618.2009.00715.x] [Medline: 19538690]

3. Yatham LN, Kennedy SH, Parikh SV, Schaffer A, Bond DJ, Frey BN, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord 2018 Mar;20(2):97-170 [FREE Full text] [doi: 10.1111/bdi.12609] [Medline: 29536616]

4. van Bendegem MA, van den Heuvel SC, Kramer LJ, Goossens PJ. Attitudes of patients with bipolar disorder toward the Life Chart Methodology: a phenomenological study. J Am Psychiatr Nurses Assoc 2014;20(6):376-385. [doi: 10.1177/1078390314558420] [Medline: 25367897]

5. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol Med 2015 Oct;45(13):2691-2704. [doi: 10.1017/S0033291715000410] [Medline: 26220802]

6. Morriss R, Faizal M, Jones A, Williamson P, Bolton C, McCarthy J. Interventions for helping people recognise early signs of recurrence in bipolar disorder. Cochrane Database Syst Rev 2007 Jan 24(1):CD004854 [FREE Full text] [doi: 10.1002/14651858.CD004854.pub2] [Medline: 17253526]

7. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016 May;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]

8. Morton E, Michalak EE, Murray G. What does quality of life refer to in bipolar disorders research? A systematic review of the construct's definition, usage and measurement. J Affect Disord 2017 Apr 1;212:128-137. [doi: 10.1016/j.jad.2017.01.026] [Medline: 28160685]

9. McIntyre RS. Understanding needs, interactions, treatment, and expectations among individuals affected by bipolar disorder or schizophrenia: the UNITE global survey. J Clin Psychiatry 2009;70(Suppl 3):5-11 [FREE Full text] [doi: 10.4088/JCP.7075su1c.02] [Medline: 19570496]

10. Eiring Ø, Nylenna M, Nytrøen K. Patient-important outcomes in the long-term treatment of bipolar disorder: a mixed-methods approach investigating relative preferences and a proposed taxonomy. Patient 2016 Apr;9(2):91-102. [doi: 10.1007/s40271-015-0128-x] [Medline: 25990222]

11. Haarig F, Berndt C, Kühnert M, Fuchs S, Bräunig P, Mühlig S. What is important to patients? Determination of patient-centered therapeutic goals in the treatment of bipolar disorders. Zeitschrift für Psychiatrie, Psychologie und Psychotherapie 2016 Mar;64(2):111-120. [doi: 10.1024/1661-4747/a000269]

12. Maczka G, Siwek M, Skalski M, Dudek D, Grabski B. Patients' and doctors' attitudes towards bipolar disorder – do we share our beliefs? Arch Psychiatry Psychother 2010;2:43-50 [FREE Full text]

13. Sajatovic M, Jenkins JH, Cassidy KA, Muzina DJ. Medication treatment perceptions, concerns and expectations among depressed individuals with Type I Bipolar Disorder. J Affect Disord 2009 Jun;115(3):360-366 [FREE Full text] [doi: 10.1016/j.jad.2008.10.002] [Medline: 18996600]

14. Michalak EE, Hole R, Livingston JD, Murray G, Parikh SV, Lapsley S, et al. Improving care and wellness in bipolar disorder: origins, evolution and future directions of a collaborative knowledge exchange network. Int J Ment Health Syst 2012 Sep 10;6(1):16 [FREE Full text] [doi: 10.1186/1752-4458-6-16] [Medline: 22963889]

15. Murray G, Leitan ND, Thomas N, Michalak EE, Johnson SL, Jones S, et al. Towards recovery-oriented psychosocial interventions for bipolar disorder: Quality of life outcomes, stage-sensitive treatments, and mindfulness mechanisms. Clin Psychol Rev 2017 Mar;52:148-163. [doi: 10.1016/j.cpr.2017.01.002] [Medline: 28129636]

16. Michalak EE, Morton E, Barnes SJ, Hole R, CREST.BD, Murray G. Supporting self-management in bipolar disorder: mixed-methods knowledge translation study. JMIR Ment Health 2019 Apr 15;6(4):e13493 [FREE Full text] [doi: 10.2196/13493] [Medline: 30985287]

17. Matsuda A, Yamada Y, Ishizuka N, Matsushima E, Kobayashi K, Ohkubo T, et al. Effects of a self-monitoring quality of life intervention for patients with cancer receiving palliative care in Japan: study protocol for a randomized controlled trial. Asian Pac J Cancer Prev 2018 Nov 28;19(11):3027-3032 [FREE Full text] [doi: 10.31557/APJCP.2018.19.11.3027] [Medline: 30484987]

18. Kobayashi K, Green J, Shimonagayoshi M, Kanemoto N, Kasai R, Itoh Y, et al. Validation of the care notebook for measuring physical, mental and life well-being of patients with cancer. Qual Life Res 2005 May;14(4):1035-1043. [doi: 10.1007/s11136-004-2958-1] [Medline: 16041899]

19. Fischer MJ, Krol-Warmerdam EM, Ranke GM, Zegers MH, Averink RA, Scholten AN, et al. Routine monitoring of quality of life for patients with breast cancer: an acceptability and field test. J Psychosoc Oncol 2012;30(2):239-259. [doi: 10.1080/07347332.2011.644398] [Medline: 22416958]

20. Faurholt-Jepsen M, Munkholm K, Frost M, Bardram JE, Kessing LV. Electronic self-monitoring of mood using IT platforms in adult patients with bipolar disorder: a systematic review of the validity and evidence. BMC Psychiatry 2016 Jan 15;16:7 [FREE Full text] [doi: 10.1186/s12888-016-0713-0] [Medline: 26769120]

21. Simon J, Budge K, Price J, Goodwin G, Geddes J. Remote mood monitoring for adults with bipolar disorder: an explorative study of compliance and impact on mental health service use and costs. Eur Psychiatry 2017 Sep;45:14-19 [FREE Full text] [doi: 10.1016/j.eurpsy.2017.06.007] [Medline: 28728090]

22. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

23. Denicoff KD, Leverich GS, Nolen WA, Rush AJ, McElroy SL, Keck PE, et al. Validation of the prospective NIMH-Life-Chart Method (NIMH-LCM-p) for longitudinal assessment of bipolar illness. Psychol Med 2000 Nov;30(6):1391-1397. [doi: 10.1017/s0033291799002810] [Medline: 11097079]

24. Bradstreet S, Allan S, Gumley A. Adverse event monitoring in mHealth for psychosis interventions provides an important opportunity for learning. J Ment Health 2019 Oct;28(5):461-466. [doi: 10.1080/09638237.2019.1630727] [Medline: 31240970]

25. 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]

26. Michalak EE, Murray G, Collaborative RESearch Team to Study Psychosocial Issues in Bipolar Disorder (CREST.BD). Development of the QoL.BD: a disorder-specific scale to assess quality of life in bipolar disorder. Bipolar Disord 2010 Nov;12(7):727-740. [doi: 10.1111/j.1399-5618.2010.00865.x] [Medline: 21040290]

27. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]

28. Michalak E, Murray G. Bipolar Wellness Centre. 2015. URL: http://www.bdwellness.com/ [accessed 2019-09-29]

29. Michalak E, Murray G. CREST.BD Quality of Life Tool. 2015. URL: https://www.bdqol.com/ [accessed 2019-09-29]

30. Michalak EE, Yatham LN, Kolesar S, Lam RW. Bipolar disorder and quality of life: a patient-centered perspective. Qual Life Res 2006 Mar;15(1):25-37. [doi: 10.1007/s11136-005-0376-7] [Medline: 16411028]

31. Kupfer DJ, Frank E, Grochocinski VJ, Cluss PA, Houck PR, Stapf DA. Demographic and clinical characteristics of individuals in a bipolar disorder case registry. J Clin Psychiatry 2002 Mar;63(2):120-125. [doi: 10.4088/jcp.v63n0206] [Medline: 11874212]

32. Patton MQ. Qualitative Research and Evaluation Methods. Third Edition. Thousand Oaks, CA: SAGE; 2002.

33. Morton E, Michalak E, Hole R, Buzwell S, Murray G. The 'new normal': relativity of quality of life judgments in individuals with bipolar disorder-a qualitative study. Qual Life Res 2018 Jun;27(6):1493-1500. [doi: 10.1007/s11136-018-1811-x] [Medline: 29435800]

34. Morton E, Michalak E, Hole R, Buzwell S, Murray G. 'Taking back the reins' - a qualitative study of the meaning and experience of self-management in bipolar disorder. J Affect Disord 2018 Mar 1;228:160-165. [doi: 10.1016/j.jad.2017.12.018] [Medline: 29248822]

35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

36. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]

37. Tsanas A, Saunders KE, Bilderbeck AC, Palmius N, Osipov M, Clifford GD, et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. J Affect Disord 2016 Nov 15;205:225-233 [FREE Full text] [doi: 10.1016/j.jad.2016.06.065] [Medline: 27449555]

38. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, del mar Bonnín C, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLe smartphone application: feasibility, acceptability and satisfaction. J Affect Disord 2016 Aug;200:58-66. [doi: 10.1016/j.jad.2016.04.042] [Medline: 27128358]

39. Bowling A. Current state of the art in quality of life measurement. In: Carr AJ, Higginson I, Robinson PG, editors. Quality of Life. London: BMJ Books; 2008:1-8.

40. Stastny P, Amering M. Consumer-interests and the quality of life concept - common ground or parallel universes? In: Katschnig H, Freeman H, Sartorius N, editors. Quality of Life in Mental Disorders. Second Edition. Chichester: Wiley; 2006:275-283.

41. McGee D, del Vento A, Bavelas JB. An interactional model of questions as therapeutic interventions. J Marital Fam Ther 2005 Oct;31(4):371-384. [doi: 10.1111/j.1752-0606.2005.tb01577.x] [Medline: 16294686]

42. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 6;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]

43. Hartmann JA, Wichers M, Menne-Lothmann C, Kramer I, Viechtbauer W, Peeters F, et al. Experience sampling-based personalized feedback and positive affect: a randomized controlled trial in depressed patients. PLoS One 2015;10(6):e0128095 [FREE Full text] [doi: 10.1371/journal.pone.0128095] [Medline: 26034983]

44. Kramer I, Simons CJ, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. World Psychiatry 2014 Mar;13(1):68-77 [FREE Full text] [doi: 10.1002/wps.20090] [Medline: 24497255]

45. Inder ML, Crowe MT, Moor S, Luty SE, Carter JD, Joyce PR. 'I actually don't know who I am': the impact of bipolar disorder on the development of self. Psychiatry 2008;71(2):123-133. [doi: 10.1521/psyc.2008.71.2.123] [Medline: 18573034]

46. Michalak E, Livingston JD, Hole R, Suto M, Hale S, Haddock C. 'It's something that I manage but it is not who I am': reflections on internalized stigma in individuals with bipolar disorder. Chronic Illn 2011 Sep;7(3):209-224. [doi: 10.1177/1742395310395959] [Medline: 21357643]

47. Matthews M, Murnane E, Snyder J. Quantifying the changeable self: the role of self-tracking in coming to terms with and managing bipolar disorder. Human–Comput Interact 2017;32(5-6):413-446. [doi: 10.1080/07370024.2017.1294983]

48. Skantze K, Malm U. A new approach to facilitation of working alliances based on patients' quality of life goals. Nord J Psychiatry 1994;48(1):37-55. [doi: 10.3109/08039489409078125]

49. Berk M, Berk L, Castle D. A collaborative approach to the treatment alliance in bipolar disorder. Bipolar Disord 2004 Dec;6(6):504-518. [doi: 10.1111/j.1399-5618.2004.00154.x] [Medline: 15541066]

50. Moore PJ, Little MA, McSharry PE, Goodwin GM, Geddes JR. Correlates of depression in bipolar disorder. Proc Biol Sci 2014 Mar 7;281(1776):20132320 [FREE Full text] [doi: 10.1098/rspb.2013.2320] [Medline: 24352942]

51. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 7;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]

52. Vaismoradi M, Jones J, Turunen H, Snelgrove S. Theme development in qualitative content analysis and thematic analysis. J Nano Electron Phys 2016;6(5). [doi: 10.5430/jnep.v6n5p100]

Abbreviations

BD: bipolar disorder
BD-I: bipolar I disorder
BD-II: bipolar II disorder
BD-NOS: bipolar disorder not otherwise specified
CREST.BD: Collaborative Research Team for Psychosocial Issues in Bipolar Disorders
LCM: Life Chart Methodology
QoL: quality of life

XSL•FO
RenderX

Review

Accuracy of Machine Learning Algorithms for the Diagnosis of Autism Spectrum Disorder: Systematic Review and Meta-Analysis of Brain Magnetic Resonance Imaging Studies

Sun Jae Moon[1], MD; Jinseub Hwang[2], PhD; Rajesh Kana[3], PhD; John Torous[4], MD, MBI; Jung Won Kim[5], MD

[1]Ewha Womans University Mokdong Hospital, Ewha Womans University Medical Center, Seoul, Republic of Korea

[2]Department of Computer Science and Statistics, Daegu University, Gyeongsangbuk-do, Republic of Korea

[3]Department of Psychology, University of Alabama at Tuscaloosa, Tuscaloosa, AL, United States

[4]Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[5]Department of Psychiatry and Behavioral Neurobiology, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Sun Jae Moon, MD
Ewha Womans University Mokdong Hospital
Ewha Womans University Medical Center
1071, Anyangcheon-ro, Yangcheon-gu
Seoul, 07985
Republic of Korea
Phone: 82 1022271081
Email: sunjaemoon16@gmail.com

Abstract

Background: In the recent years, machine learning algorithms have been more widely and increasingly applied in biomedical fields. In particular, their application has been drawing more attention in the field of psychiatry, for instance, as diagnostic tests/tools for autism spectrum disorder (ASD). However, given their complexity and potential clinical implications, there is an ongoing need for further research on their accuracy.

Objective: This study aimed to perform a systematic review and meta-analysis to summarize the available evidence for the accuracy of machine learning algorithms in diagnosing ASD.

Methods: The following databases were searched on November 28, 2018: MEDLINE, EMBASE, CINAHL Complete (with Open Dissertations), PsycINFO, and Institute of Electrical and Electronics Engineers Xplore Digital Library. Studies that used a machine learning algorithm partially or fully for distinguishing individuals with ASD from control subjects and provided accuracy measures were included in our analysis. The bivariate random effects model was applied to the pooled data in a meta-analysis. A subgroup analysis was used to investigate and resolve the source of heterogeneity between studies. True-positive, false-positive, false-negative, and true-negative values from individual studies were used to calculate the pooled sensitivity and specificity values, draw Summary Receiver Operating Characteristics curves, and obtain the area under the curve (AUC) and partial AUC (pAUC).

Results: A total of 43 studies were included for the final analysis, of which a meta-analysis was performed on 40 studies (53 samples with 12,128 participants). A structural magnetic resonance imaging (sMRI) subgroup meta-analysis (12 samples with 1776 participants) showed a sensitivity of 0.83 (95% CI 0.76-0.89), a specificity of 0.84 (95% CI 0.74-0.91), and AUC/pAUC of 0.90/0.83. A functional magnetic resonance imaging/deep neural network subgroup meta-analysis (5 samples with 1345 participants) showed a sensitivity of 0.69 (95% CI 0.62-0.75), specificity of 0.66 (95% CI 0.61-0.70), and AUC/pAUC of 0.71/0.67.

Conclusions: The accuracy of machine learning algorithms for diagnosis of ASD was considered acceptable by few accuracy measures only in cases of sMRI use; however, given the many limitations indicated in our study, further well-designed studies are warranted to extend the potential use of machine learning algorithms to clinical settings.

Trial Registration: PROSPERO CRD42018117779; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=117779

Introduction

Background

Autism spectrum disorder (ASD), behaviorally characterized by a deficit in social communication and rigidity in interest or behavior by both the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) and the International Statistical Classification of Diseases-11 (ICD-11), is believed to be a product of complex interactions between genetic and environmental factors [1-3]. The latest prevalence of ASD has been reported to be 1 in 59 children aged 8 years, based on the 2014 Center for Disease Control and Prevention (CDC) surveillance data [4], and 1 in 40 children aged 3-17 years, based on parental reports of the diagnosis in a national survey [5]. Despite the advancement of many biomarkers with potential in prediction or early detection of ASD (eg, structural magnetic resonance imaging [sMRI] or functional magnetic resonance imaging [fMRI]), a diagnosis is not made until the age of 4-5 years, on average [4,6].

Machine learning has been increasingly studied as a novel tool to enhance the accuracy of diagnosis and early detection of ASD [7]. Unlike traditional rule-based algorithms that allowed computers to generate answers with preprogramed rules, machine learning allows building of an algorithm that can learn, predict, and improve with experience, based on big data [3,8-10]. Psychiatric decision making is more sophisticated and difficult to characterize, compared with machine learning, although there are some common elements. Psychiatrists diagnose patients by observing their behaviors and registering all collected and collateral data into their (psychiatrists') cognitive system as sensory input values (eg, voice and vision). Similarly, machine learning requires a series of steps, including preprocessing (eg, noise removal from data before input into an algorithm), segmentation, and feature extraction [7]. In particular, machine learning in the field of ASD diagnostics incorporates big data (eg, neuroimaging), making the input data immense and complex [11]. The application of machine learning algorithms in the field of neuroimaging often requires an extra process, such as *feature selection* that extracts key features from a complex dataset. In other words, key features are selected before the learning process, which is called *feature selection* [11].

Objective

Currently, machine learning is widely applied to the field of bioinformatics, including genetics and imaging, and many applications require signal recognition and processing [12]. Machine learning algorithms are currently applied to the field of psychiatry in areas such as genomics, electroencephalogram (EEG), and neuroimaging. However, owing to the complex workflows implicated in machine learning itself, the accuracy of such algorithms is varied [8]. This study aimed to suggest an integrated estimate of the accuracy for use of machine learning algorithms in distinguishing individuals with ASD from control groups through systematic review and meta-analysis of the available studies.

Methods

Systematic Review

This systematic review and meta-analysis was conducted based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic test accuracy [13]. The study protocol was written before initiation of the study and registered in the Prospective Register of Systematic Reviews database (trial registration: CRD42018117779).

Data Sources and Search Strategy

MEDLINE, EMBASE, CINAHL Complete (with Open Dissertations), and PsycINFO were selected as core search databases, and the Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library was added to maximize the sensitivity of the search. The IEEE Xplore Digital Library is a database created by the IEEE, the largest of its kind worldwide, and includes more than 1800 peer-reviewed conference proceedings. Default search filters provided by journals were not used. There was no restriction by publication type (eg, conference proceedings) or language. The initial search was conducted on November 28, 2018. The search strategy and query per search database are listed in Multimedia Appendix 1. The primary consideration for study inclusion was if machine learning was partially or fully applied in distinguishing individuals clinically diagnosed with ASD from controls and assess the accuracy of such applications. Multimedia Appendix 2 lists inclusion/exclusion criteria. An author (SM) retrieved the initial search results and removed duplicates by using the command *find duplicate* via a reference software (Endnote X9, Clarivate Analytics, Philadelphia, Pennsylvania. Subsequently, another author (JK) manually searched for and removed any residual duplicates. Finally, the studies were screened independently by two authors (SM and JK) by title, abstract, and keywords, after which the full texts of the selected studies were screened by two authors (SM and JK) by inclusion/exclusion criteria. If any discrepancy was found in the final selection, the two authors reached a consensus via discussion.

Data Extraction

A data extraction form was created through discussion among the authors before the extraction process to suggest specific subgroups and coding processes (categorizing) for a meta-analysis (Multimedia Appendix 3). The process is provided in detail in Multimedia Appendix 4. General characteristics such as author, publication year, sample size, average age, gender ratio, and data characteristics were extracted from individual studies. Information regarding the reference standard used in individual studies and definitions of positive/negative disease (autism positive/control) and methodologies to distinguish individuals with autism from control group were collected. Specific methodologies used to process and classify data for use in machine learning algorithms were also recorded (Multimedia Appendices 3 and 4). All accuracy values were extracted, and true-positive / true-negative / false-positive /

false-negative (TP/TN/FP/FN) values were calculated from individual studies for a meta-analysis. If the TP/TN/FP/FN values could not be calculated from the accuracy values provided in a study, an email was sent to the corresponding author to request raw data. If there was no response within 14 days, the study was not included in the meta-analysis. The extraction was performed independently by two authors (SM and JK). If there was any discrepancy in the extracted data, a consensus was reached by thorough discussion after repeating the same extraction process.

Quality Assessment

Two authors (SM and JK) independently assessed the quality of individual studies based on the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). QUADAS-2 is a validated tool used to evaluate the quality of diagnostic accuracy studies by patient selection, index test, reference standard, and risk of bias (RoB) for internal validity and external validity for applicability concerns of individual studies [14]. There was no disagreement between authors in the assessment of patient selection and reference standard domain. The index test, also known as the target tool of our investigation in this study, is a machine learning algorithm. The target tool, the machine learning algorithm's accuracy, is reported through a process called validation. However, when a study provided no information about the validation process, low RoB was assumed if independent datasets were used for training, building a model, and validation [15]. Otherwise, the level of RoB was determined by thoroughly reviewing the validation processes.

Evidence Synthesis

In our meta-analysis, a bivariate random effects model was used to consider both within- and between-subject variability and threshold effect [16]. A Summary Receiver Operating Characteristics (SROC) curve was generated based on parameter estimates extracted from the bivariate random effects model [17]. The SROC curve was specified by pooled sensitivity, specificity point, 95% CIs, and prediction region. Area under the curve (AUC) and partial AUC (pAUC) were calculated based on the SROC curve [18]. Studies that were visually deviant from the 95% prediction region on the SROC curve were considered heterogeneous [19]. Attempts were made to resolve the heterogeneity by performing a subgroup analysis—generating individual SROC curves for subgroups (minimum 5 studies) [20]. If most studies were within the 95% prediction region on the SROC curves of the subgroups, the sample was determined to be homogeneous, and integrated sensitivity, specificity, and SROC curve results were provided.

If any of the TP/FP/TN/FN value was 0, 0.5 was added to prevent zero cell count problem [21]. The TP/FP/TN/FN values were extracted or calculated from each independent sample in a study, and if multiple machine learning algorithms were applied to the same sample, an algorithm with the best accuracy (calculated as [TP+TN]/[TP+FP+TN+FN]) was selected for data extraction.

A meta-analysis was conducted via the mada package in R (version 3.4.3, R Core Team, Vienna, Austria), and statistical significance was expressed with 95% CIs. Publication bias was not assessed in our analysis, as there are currently no statistically adequate models in the field of meta-analysis of diagnostic test accuracy [22].

Results

Search, Selection, and General Characteristics

After duplicate removal, of the 280 studies extracted from five databases and one additional database, 43 studies were selected, of which 40 studies were included in the meta-analysis. Figure 1 provides details according to the screening stage.

The publication years ranged from 2007 to 2018 for the final selection of 43 studies, of which 40 were journal articles and 3 were gray literature elements (eg, conference proceedings). A total of 10 studies used a public database that was available on the internet and open to anyone, 18 used a private sector database (eg, clinic and hospital), 3 used both public and private databases, and the remaining 12 used databases from others. Regarding the average age of the sample, 5 studies included adults, 22 studies included school-aged participants, 11 included preschool-aged participants, and the remaining 5 did not provide any information. For the machine learning algorithm, 20 studies used a support vector machine (SVM), 3 used a deep neural network (DNN), 13 used others, and the remaining 10 used and compared multiple algorithms. For prediction, 11 studies used sMRI features, 9 used fMRI features, 9 used behavior traits, 5 used biochemical features, 4 used EEG features, and the remaining 2 used text or voice features. For reference standards, 24 studies used DSM-IV, DSM-IV - Text Revision, or DSM-5; 10 used the Autism Diagnostic Observation Schedule (ADOS) or the Autism Diagnostic Interview (ADI); 2 used ICD; and the remaining 7 did not provide relevant information. For the validation methodology, 37 studies only used internal validation, 2 only used external validation, and 4 used both. The abovementioned information is summarized in Table 1, and the extracted raw data are presented in Multimedia Appendices 5 and 6.
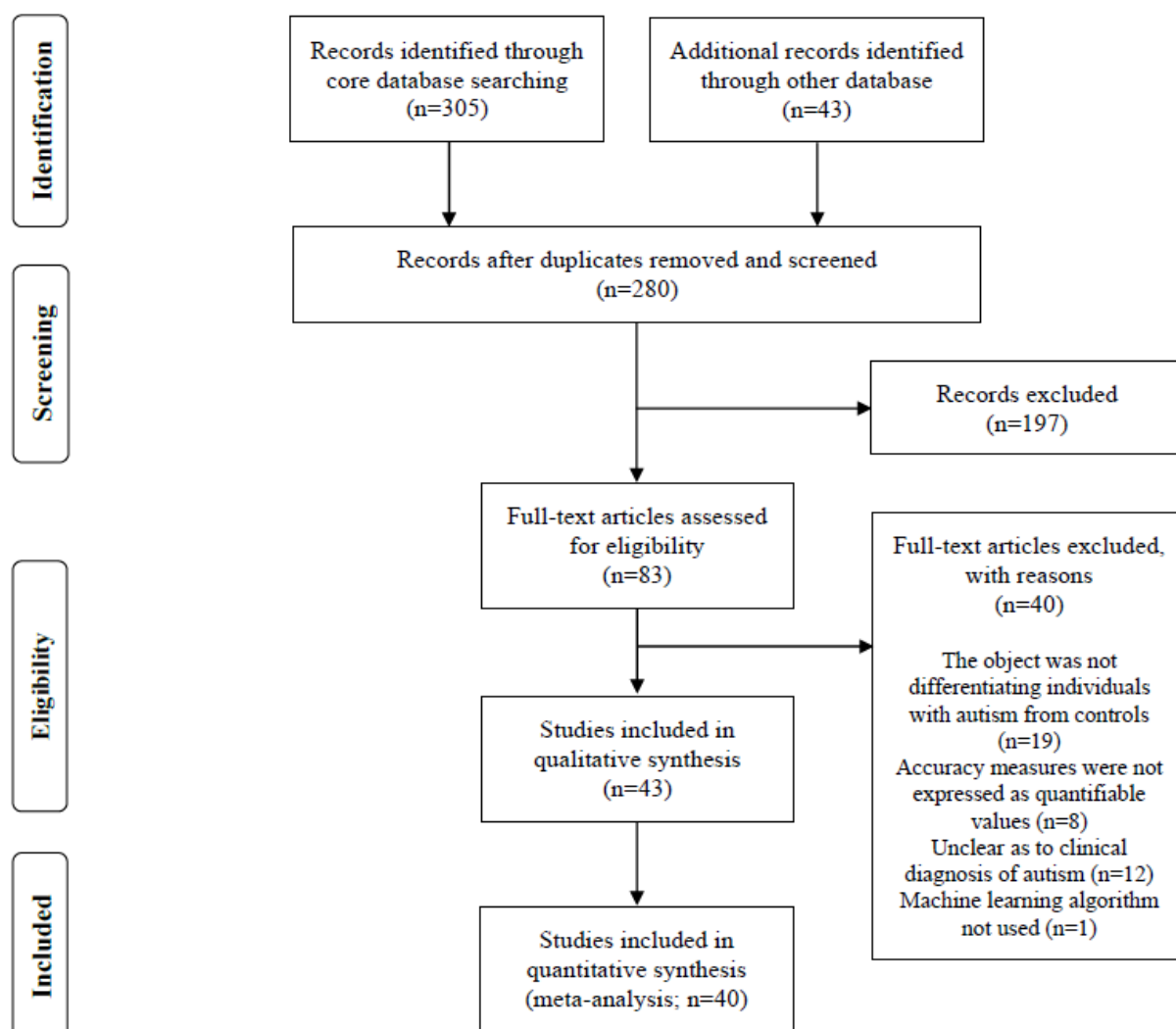
Figure 1. Flowchart for the literature screening and selection process.

Table 1. Characteristics of 43 studies for the systematic review and 53 samples for the meta-analysis.

Characteristics	Studies (n)[a]	Samples (n)[b]
Publication type		
Journal article	40	50
Letter, report, or conference proceeding	3	3
Dataset type		
Private (hospital or clinic) dataset	18	21
Public database	10	16
Mixed (private and public) dataset	3	0
Others or unknown	12	16
Mean age of sample (years)		
Adults (≥18)	5	5
School age (6-18)	22	27
Preschool age (<6)	11	16
Unknown	5	5
Classification algorithm type		
Support vector machine	20	24
Deep neural network	3	6
Others[c]	13	23
Mixed	10	0
Predictor type		
Structural MRI[d] features	11	14
Functional MRI features[e]	9	13
Behavior traits	9	14
Biochemical features	5	7
Electroencephalography features	4	3
Text or voice	2	2
Reference standard		
DSM[f]-IV (Text Revision) or DSM-5	24	28
ADOS[g] or ADI[h]	10	12
ICD[i]	2	2
Others or not otherwise specified	7	11
Validation method		
Internal validation	36	46
External validation	2	6
Internal and external validation	4	0
Others or not otherwise specified	1	1

[a]Number of studies for a given category (N=43 in total).

[b]Number of datasets used in studies (N=53 in total).

[c]Probabilistic neural network, decision tree, regression, ensemble, random forest, and fuzzy.

[d]MRI: magnetic resonance imaging.

[e]All studies used resting-state MRI images (one study used both resting state and task-related MRI images).

[f]DSM: Diagnostic and Statistical Manual of Mental Disorders.

[g]ADOS: Autism Diagnostic Observation Schedule.

XSL•FO

RenderX

[h]ADI: Autism Diagnostic Interview.

[i]ICD: International Statistical Classification of Diseases.

Qualitative Assessment

Of the 43 studies in total, more than half were assessed to have an unclear RoB by patient selection domain (33 studies) and index test domain (29 studies). More than half were considered to have a low RoB by the total reference standard (35 studies) and flow and timing domains (35 studies). For applicability concern, about half (22 studies) were shown to have unclear or high-risk RoB by patient selection domain, whereas most were considered to have a low risk by index test (42 studies) and reference standard domain (36 studies). Qualitative assessment for all the individual studies is summarized in Multimedia Appendix 7, and the distribution is shown in Figure 2.

Figure 2. Risk of bias and applicability concern by domain in Quality Assessment of Diagnostic Accuracy Studies-2. Microsoft Excel was used.
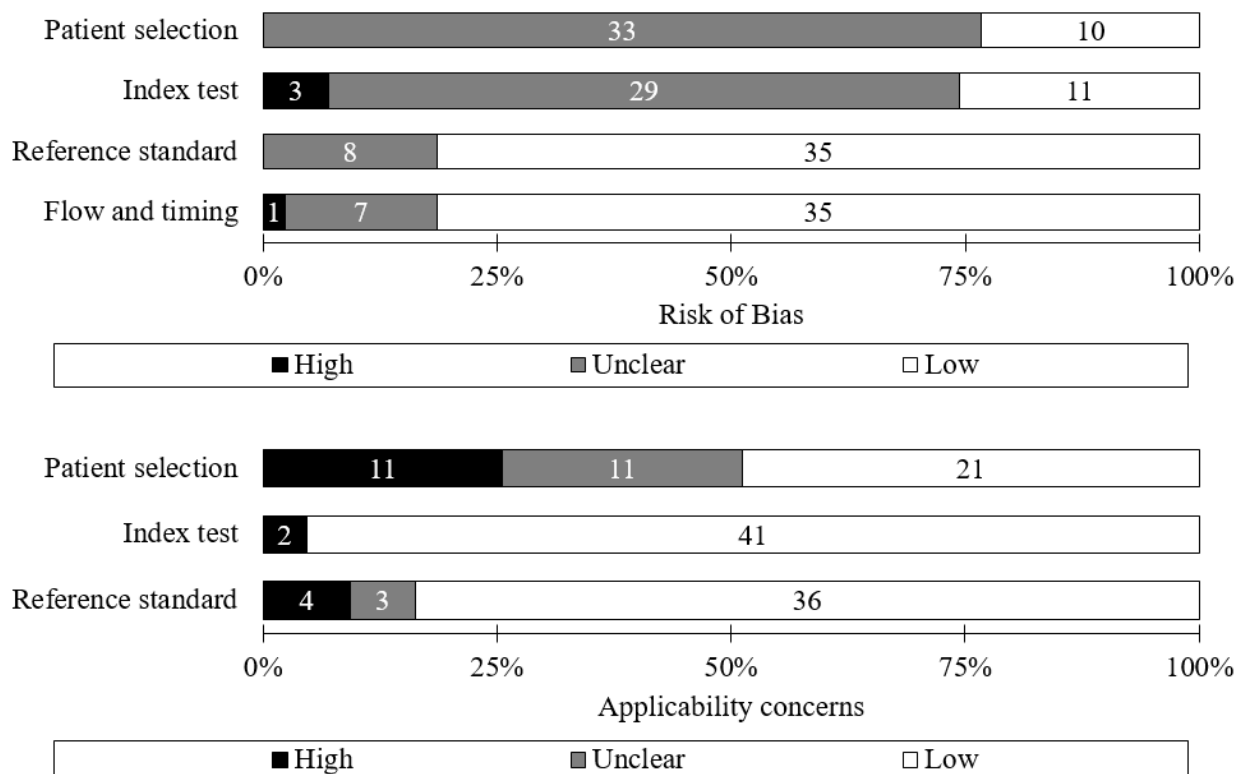
Quantitative Analysis (Meta-Analysis)

Of the final selection of 43 studies, only 40, from which TP/FP/FN/TN values were extractable, were considered for the meta-analysis. A total of 53 independent samples were extracted from the 40 studies and included in the meta-analysis (Table 1). Of the 53 samples, 12,128 participants were inspected in the meta-analysis, with the total sensitivity and specificity ranging from 0.55 to 1.00 and 0.56 to 0.99, respectively. TP/FP/FN/TN, sensitivity, and specificity values for 53 individual samples are summarized in Multimedia Appendix 8, and visual distribution is provided as SROC in Figure 3. Of the 53 samples, 12 were found outside the 95% predictive region of the SROC curve, and therefore, there was heterogeneity between samples (Figure 3).

In an attempt to resolve this heterogeneity, a subgroup analysis was conducted with 19 variables that had been predefined and coded. For replicability, a raw data sheet listing the precodified variables is available in Multimedia Appendix 9. As a result, among 19 variables, *predictor* was the only one by which the heterogeneity could be partially resolved. Of the 53 samples,

for the sMRI subgroup that used sMRI as predictors, all the 12 samples were found to be within the predictive region of the SROC curve, thus resolving the heterogeneity (Figure 4).

For the sMRI subgroup, the pooled sensitivity was 0.83 (95% CI 0.76-0.89), specificity was 0.84 (95% CI 0.74-0.91), and AUC/pAUC was 0.90/0.83. Meta-analysis was also attempted for the remaining subgroups, such as fMRI (15 samples), behavior traits (14 samples), and biochemical features (7 samples) subgroups, but the pooled sensitivity and specificity could not be provided owing to a significant degree of heterogeneity between samples: A few samples were shown to be far off the predictive region of the SROC curves (Multimedia Appendices 10-12). However, sub-subgroup meta-analysis using 5 samples that used fMRI as a predictor and DNN as a classifier allowed for the heterogeneity to be resolved and provided the pooled sensitivity of 0.69 (95% CI 0.62-0.75), specificity of 0.66 (95% CI 0.61-0.70), and AUC/pAUC of 0.71/0.67 (Figure 5).

Similarly, another sub-subgroup meta-analysis of six samples that used sMRI as a predictor and SVM as a classifier resolved

the heterogeneity and resulted in a pooled sensitivity of 0.87 (95% CI 0.78-0.93), specificity of 0.87 (95% CI 0.71-0.95), and AUC/pAUC of 0.92/0.88 (Multimedia Appendix 12). Sensitivity and specificity values and types of classifiers used for samples of individual subgroups that used neuroimaging features (sMRI and fMRI subgroups) as predictors are provided in Table 2, and a forest plot is provided in Multimedia Appendix 13.

Summary Receiver Operating Characteristics curve for functional magnetic resonance imaging/deep neural network sub-subgroup (5 samples). Note that confidence region is the 95% confidence region around the summary sensitivity and specificity points, and the prediction region is the 95% prediction of the true sensitivity and specificity interval for

future observations. SROC: Summary Receiver Operating Characteristics.

The sensitivity and specificity for the behavior traits (14 samples) subgroup ranged from 0.68 to 1.00 and 0.56 to 0.9, respectively. The sensitivity and specificity for the biochemical features (7 samples) subgroup ranged from 0.77 to 0.94 and 0.72 to 0.93, respectively. The sensitivity and specificity for the EEG subgroup (3 samples) ranged from 0.94 to 0.97 and 0.81 to 0.94, respectively. The results are summarized in Multimedia Appendix 8. Information for other measures not included in the meta-analysis is provided in Multimedia Appendix 14.

Figure 3. Summary Receiver Operating Characteristics curve for all 53 samples. Note that the confidence region is the 95% confidence region around the summary sensitivity and specificity points, and the prediction region is the 95% prediction of the true sensitivity and specificity interval for future observations. SROC: Summary Receiver Operating Characteristics.
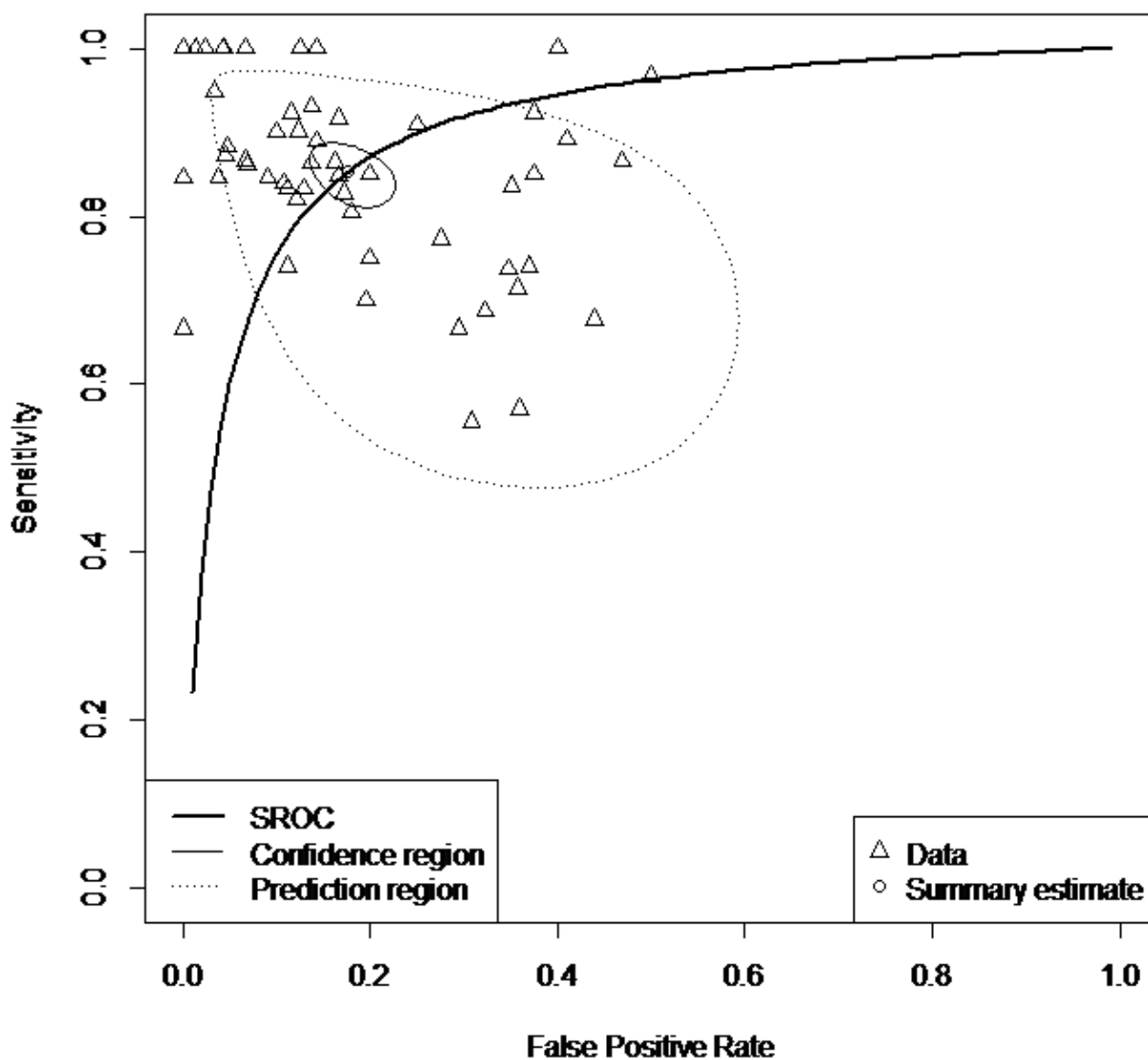
Figure 4. Summary Receiver Operating Characteristics curve for structural magnetic resonance imaging subgroup (12 samples). Note that the confidence region is the 95% confidence region around the summary sensitivity and specificity points, and the prediction region is the 95% prediction of the true sensitivity and specificity interval for future observations. SROC: Summary Receiver Operating Characteristics.
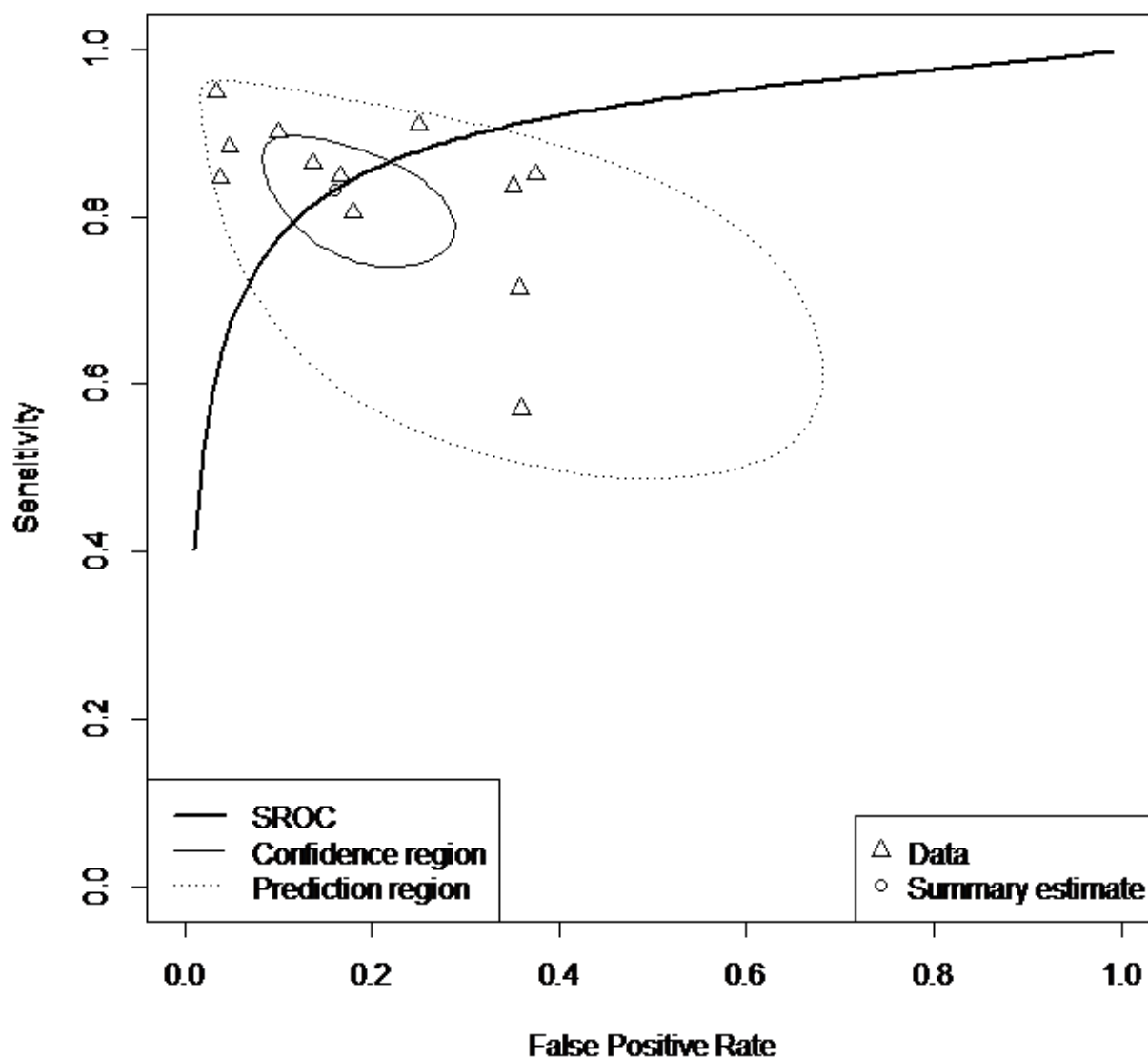
Figure 5.
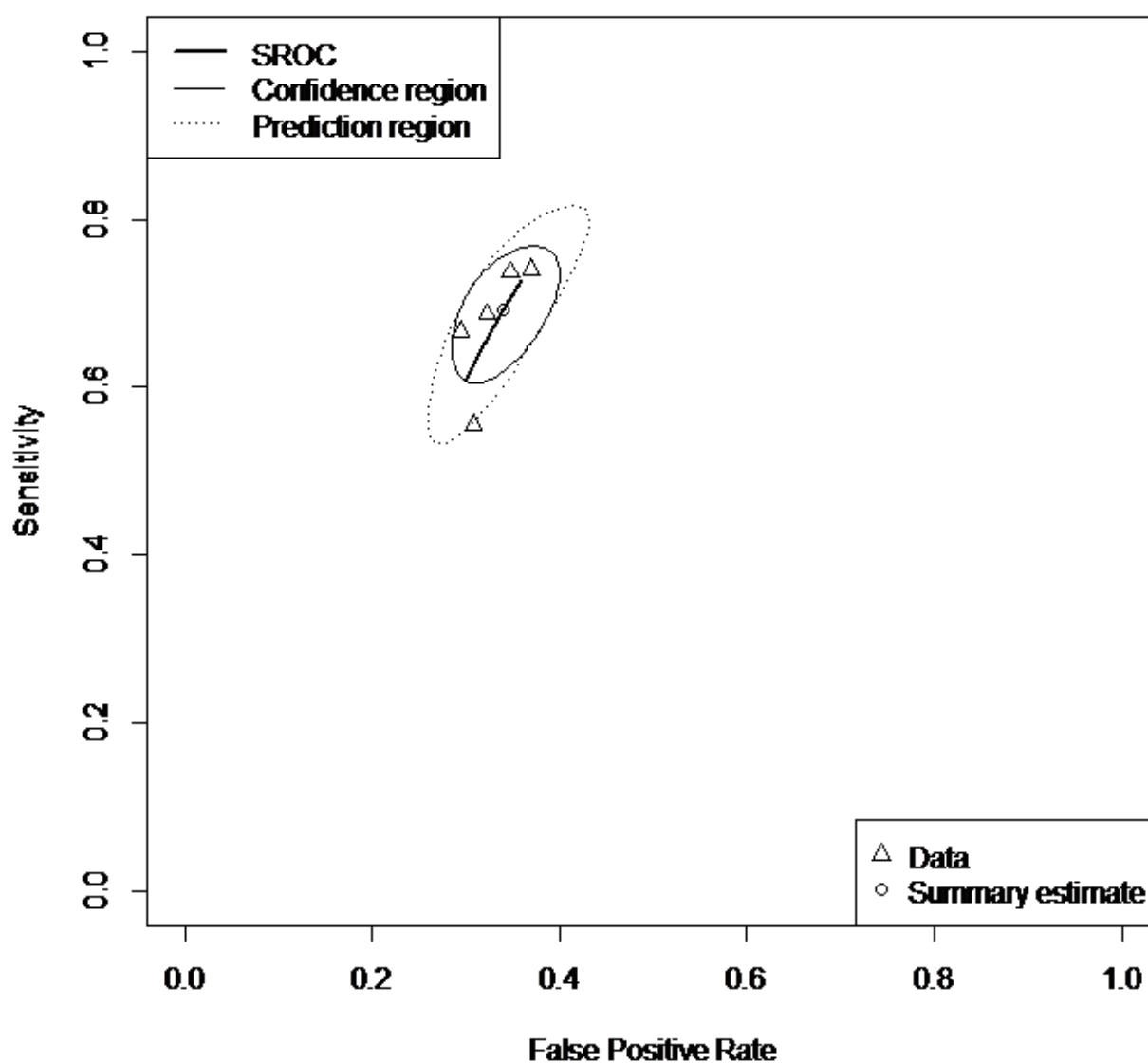
Table 2. Sensitivity and specificity of structural and functional magnetic resonance imaging subgroups.

MRI[a] group	Sample size (n)	Sensitivity (95% CI)	Specificity (95% CI)
Structural MRI subgroup			
Hazlett et al (2017) [23]	179	0.87 (0.72-0.95)	0.95 (0.90-0.97)
Chaddad et al (2017)[b] [24]	28	0.70 (0.45-0.87)	0.63 (0.39-0.83)
Chaddad et al (2017)[c] [24]	36	0.83 (0.63-0.94)	0.62 (0.39-0.81)
Wee et al (2014) [25]	117	0.94 (0.85-0.98)	0.96 (0.88-0.99)
Ecker et al (2010)[d] [26]	44	0.85 (0.65-0.94)	0.85 (0.65-0.94)
Ecker et al (2010)[d] [27]	40	0.88 (0.68-0.96)	0.88 (0.68-0.96)
Xiao et al (2017) [28]	85	0.80 (0.66-0.89)	0.81 (0.67-0.90)
Katuwal et al (2015) [29]	734	0.57 (0.52-0.62)	0.64 (0.59-0.69)
Jiao et al (2010) [30]	38	0.89 (0.71-0.97)	0.74 (0.50-0.89)
Neeley et al (2007) [31]	57	0.84 (0.68-0.93)	0.82 (0.63-0.92)
Kong et al (2019) [32]	182	0.84 (0.75-0.91)	0.96 (0.90-0.98)
Shen et al (2018) [33]	236	0.83 (0.77-0.88)	0.65 (0.54-0.74)
Subtotal by range and pooled estimate from meta-analysis	1776	0.57-0.94; 0.83 (0.76-0.89)	0.62-0.96; 0.84 (0.74-0.91)
Functional MRI subgroup			
Li et al (2018)[b] [34]	113	0.68 (0.54-0.80)	0.67 (0.55-0.78)
Li et al (2018)[e] [34]	75	0.55 (0.40-0.70)	0.69 (0.53-0.81)
Li et al (2018)[f] [34]	61	0.73 (0.58-0.84)	0.65 (0.45-0.80)
Li et al (2018)[g] [34]	61	0.66 (0.48-0.81)	0.70 (0.54-0.83)
Heinsfeld et al (2018) [35]	1035	0.74 (0.70-0.78)	0.63 (0.59-0.67)
Dekhil et al (2018) [36]	283	0.90 (0.83-0.94)	0.88 (0.82-0.92)
Bernas et al (2018)[g] [37]	30	0.89 (0.62-0.97)	0.81 (0.54-0.94)
Mastrovito et al (2018) [38]	54	0.73 (0.55-0.86)	0.88 (0.71-0.95)
Emerson et al (2017) [39]	59	0.82 (0.56-0.94)	0.99 (0.91-1.00)
Price et al (2014) [40]	60	0.86 (0.69-0.94)	0.92 (0.77-0.98)
Uddin et al (2013)[h] [41]	40	0.74 (0.53-0.88)	0.79 (0.57-0.91)
Uddin et al (2013)[i] [41]	30	0.66 (0.42-0.84)	0.97 (0.76-1.00)
Wang et al (2012) [42]	58	0.82 (0.65-0.92)	0.82 (0.65-0.92)
Bernas et al (2018)[i] [37]	24	0.81 (0.54-0.94)	0.87 (0.66-0.96)
Lidaka (2015) [43]	640	0.92 (0.89-0.95)	0.88 (0.84-0.91)
Subtotal	2623	0.55-0.92	0.63-0.99
Overall (sMRI[j]+fMRI[k])	4399	0.55-0.94	0.62-0.99

[a]MRI: magnetic resonance imaging.

[b]Autism Brain Imaging Data Exchange-University of Michigan sample.

[c]Autism Brain Imaging Data Exchange-University of Pittsburgh sample.

[d]Same author years but different (independent) studies.

[e]Autism Brain Imaging Data Exchange-University of California Los Angeles sample.

[f]Autism Brain Imaging Data Exchange-University of Utah School of Medicine.

[g]Autism Brain Imaging Data Exchange-Katholieke Universiteit Leuven.

[h]National Database for Autism Research sample.

[i]Clinic sample.

[j]sMRI: structural magnetic resonance imaging.

[k]fMRI: functional magnetic resonance imaging.

Discussion

Principal Findings

On the basis of the meta-analysis in this study, the summary sensitivity and specificity of the accuracy for use of machine learning algorithms in ASD diagnosis are 0.83 (95% CI 0.76-0.89) and 0.84 (0.74-0.91), respectively, whereas the accuracy value based on AUC/pAUC is 0.90/0.83. On the basis of the opinion that the AUC/pAUC value is considered acceptable when above 0.7, both the AUC/pAUC values can be thought to be acceptable for the sMRI subgroup [44]. However, given the wide confidence interval for each summary sensitivity and specificity, the clinical usefulness of those values can be difficult to determine. In addition, precaution is warranted for interpreting the accuracy results, as the 95% predictive region is larger than the 95% CI region on the SROC curve, indicating a high degree of uncertainty for the pooled sensitivity and specificity calculated [19]. In addition, only one sample from the sMRI subgroup utilized an external validation method, where demographic characteristics of the training dataset were independent of those of the validation dataset. In other words, the rest of the samples in the sMRI subgroup built their validation datasets from participants who were similar to or the same as those recruited in the training datasets. Hence, those samples are believed to have high risks of overfitting, compromising the generalizability of machine learning models and overestimating the results of the meta-analysis of the sMRI subgroup [15].

Machine learning algorithms can be divided into supervised, unsupervised, or reinforcement learning by learning pattern [9]. SVM, for which subgroup analysis was performed for sMRI, is the oldest method of supervised learning, whereas DNN, for which subgroup analysis was conducted for fMRI, is the most advanced of the neural network methods (supervised learning), modeled after the mechanism of neurons [9]. On the contrary, the accuracy values for the fMRI subgroup using one of the latest machine learning algorithms, DNN, were found to be lower than those for the sMRI subgroup. This may, in part, be attributable to possible overestimation secondary to the overfitting in the sMRI subgroup. In addition, one of the studies in the fMRI/DNN sub-subgroup composed their dataset by recruiting over 1000 participants from various sites to minimize limitations such as overfitting in their analysis.

Limitations

Our study has several limitations. Of the final selection of 43 studies, 33 did not provide clear information regarding the process of obtaining an original database or a recruiting training/validation dataset from the real clinical world, or raw data such as basic demographic characteristics of the participants before the input process, thus increasing the RoB in the patient selection processes. For example, more than half the finally selected studies did not match the samples for age or gender, and the number of images or signals per participant was not specified in most of the neuroimaging and EEG studies. Subgroups other than the sMRI subgroup included studies that used the same database, thus raising concerns for possible sample overlap, which was challenging to process statistically owing to the lack or absence of information on the patient selection process. If datasets overlapped and lowered the accuracy, the subgroup meta-analysis would have been underestimated and vice versa. In addition, behavior, EEG, and voice/text subgroups did not consist of enough studies to attempt to resolve the heterogeneity and provide pooled accuracy values. Furthermore, owing to the heterogeneity, summary accuracy values could not be obtained for adult (aged over 18 years), school-age (between 6 and 18 years), and preschool-age (less than 6 years) subgroups, thus limiting the ability to draw a conclusion on accuracy by age groups. Corresponding authors for individual studies with small and high T*P* values (ie, 100% accurate machine learning test) were reached out to, and one responded. Even if more had responded, to our knowledge, there would not have been any way to perform the aggregation.

Comparison With Prior Work

To our knowledge, there is currently no study that has performed a systematic review and/or a meta-analysis on diagnostic test accuracy for the use of machine learning in diagnosing ASD and suggested its pooled estimate accuracies. In this analysis, many individual studies reported small TP and high TP (ie, 100% accurate machine learning test) and caused significant heterogeneity for a meta-analysis (see Figure 3). Authors resolved the heterogeneity by using subgroup analyses. As a result, individual studies with small and high T*P* values (ie, 100% accurate machine learning test) were barely included in fMRI and sMRI subgroup analyses, thereby resolving the heterogeneity and allowing conduct of the meta-analysis. Nevertheless, recommendations from our results may improve the quality of prospective studies using machine learning algorithms in ASD diagnosis. First, Standards for Reporting of Diagnostic Accuracy Studies (STARD) can guide machine learning diagnostic studies to enhance the reporting of patient selection processes. In addition, there is the comprehensive guideline for algorithm developers in terms of choosing an adequate predictive model for a target sample; setting the parameters, definition, or threshold; and minimizing errors such as overfitting and perfect separation [45]. Use of the STARD and other guidelines [45] would facilitate more transparent and comprehensive work in this space. Although not discussed in the studies included in our analysis, decision or running time for a machine learning algorithm in ASD diagnosis could become an important quality measure in the near future when these algorithms might be employed in a busy daily clinical practice.

Conclusions

The accuracy of diagnosing ASD by machine learning algorithms was found to be acceptable by select accuracy measures only in studies that utilized sMRI. However, because of the high heterogeneity in the analyzed studies, it is impossible

XSL•FO

RenderX

to draw a conclusion on any subgroups that used behavior traits or biochemical markers as predictors. There is a clear need for new studies with more comprehensive reporting of the selection process and dataset to draw a more accurate conclusion.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and results.
[PDF File (Adobe PDF File), 464 KB - mental_v6i12e14108_app1.pdf]

Multimedia Appendix 2
Inclusion and exclusion criteria.
[PDF File (Adobe PDF File), 65 KB - mental_v6i12e14108_app2.pdf]

Multimedia Appendix 3
Data extraction form and detailed information for coding of subgroups.
[PDF File (Adobe PDF File), 77 KB - mental_v6i12e14108_app3.pdf]

Multimedia Appendix 4
Detailed process of data extraction.
[DOCX File , 16 KB - mental_v6i12e14108_app4.docx]

Multimedia Appendix 5
General characteristics of studies (details).
[PDF File (Adobe PDF File), 104 KB - mental_v6i12e14108_app5.pdf]

Multimedia Appendix 6
Characteristics of the performance and validation condition (details).
[PDF File (Adobe PDF File), 132 KB - mental_v6i12e14108_app6.pdf]

Multimedia Appendix 7
Quality Assessment of Diagnostic Accuracy Studies‐2 assessment of all studies.
[PDF File (Adobe PDF File), 70 KB - mental_v6i12e14108_app7.pdf]

Multimedia Appendix 8
True positive, false positive, false negative, true negative, sensitivity, and specificity of 53 study samples.
[PDF File (Adobe PDF File), 86 KB - mental_v6i12e14108_app8.pdf]

Multimedia Appendix 9
Precoded raw variable data sheet for subgroup analysis.
[PDF File (Adobe PDF File), 106 KB - mental_v6i12e14108_app9.pdf]

Multimedia Appendix 10
Summary Receiver Operating Characteristics curve for biochemical features subgroup (7 samples).
[PDF File (Adobe PDF File), 81 KB - mental_v6i12e14108_app10.pdf]

Multimedia Appendix 11
Summary Receiver Operating Characteristics curve for behavior trait subgroup (14 samples).

[PDF File (Adobe PDF File), 84 KB - mental_v6i12e14108_app11.pdf]

Multimedia Appendix 12

Summary Receiver Operating Characteristics curve for structural magnetic resonance imaging/support vector machine sub-subgroup (6 samples).

[PDF File (Adobe PDF File), 76 KB - mental_v6i12e14108_app12.pdf]

Multimedia Appendix 13

Forest plot of structural and functional magnetic resonance imaging subgroup.

[PDF File (Adobe PDF File), 319 KB - mental_v6i12e14108_app13.pdf]

Multimedia Appendix 14

Excluded accuracy indices for meta-analysis.

[PDF File (Adobe PDF File), 395 KB - mental_v6i12e14108_app14.pdf]

References

1. Muhle RA, Reed HE, Stratigos KA, Veenstra-VanderWeele J. The emerging clinical neuroscience of autism spectrum disorder: a review. JAMA Psychiatry 2018 May 1;75(5):514-523. [doi: 10.1001/jamapsychiatry.2017.4685] [Medline: 29590280]

2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5. Arlington, VA: American Psychiatric Publishing; 2013.

3. ICD-11. 2018. ICD-11 for Mortality and Morbidity Statistics URL: https://icd.who.int/browse11/l-m/en [accessed 2019-10-21]

4. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveill Summ 2018 Apr 27;67(6):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6706a1] [Medline: 29701730]

5. Kogan MD, Vladutiu CJ, Schieve LA, Ghandour RM, Blumberg SJ, Zablotsky B, et al. The prevalence of parent-reported autism spectrum disorder among US children. Pediatrics 2018 Dec;142(6). [doi: 10.1542/peds.2017-4161] [Medline: 30478241]

6. Zwaigenbaum L, Penner M. Autism spectrum disorder: advances in diagnosis and evaluation. Br Med J 2018 May 21;361:k1674. [doi: 10.1136/bmj.k1674] [Medline: 29784657]

7. Thabtah F. Machine learning in autistic spectrum disorder behavioral research: a review and ways forward. Inform Health Soc Care 2019 Sep;44(3):278-297. [doi: 10.1080/17538157.2017.1399132] [Medline: 29436887]

8. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Mar;3(3):223-230. [doi: 10.1016/j.bpsc.2017.11.007] [Medline: 29486863]

9. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al. Current applications and future impact of machine learning in radiology. Radiology 2018 Aug;288(2):318-328 [FREE Full text] [doi: 10.1148/radiol.2018171820] [Medline: 29944078]

10. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]

11. Kassraian-Fard P, Matthis C, Balsters JH, Maathuis MH, Wenderoth N. Promises, pitfalls, and basic guidelines for applying machine learning classifiers to psychiatric imaging data, with autism as an example. Front Psychiatry 2016;7:177 [FREE Full text] [doi: 10.3389/fpsyt.2016.00177] [Medline: 27990125]

12. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]

13. McInnes MD, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the PRISMA-DTA Group, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. J Am Med Assoc 2018 Jan 23;319(4):388-396. [doi: 10.1001/jama.2017.19163] [Medline: 29362800]

14. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]

15. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 2018 Mar;286(3):800-809. [doi: 10.1148/radiol.2017171920] [Medline: 29309734]

16. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005 Oct;58(10):982-990. [doi: 10.1016/j.jclinepi.2005.02.022] [Medline: 16168343]

17. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MG, Stijnen T. Bivariate random effects meta-analysis of ROC curves. Med Decis Making 2008;28(5):621-638. [doi: 10.1177/0272989X08319957] [Medline: 18591542]

18. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg 2005 Jan;79(1):16-20. [doi: 10.1016/j.athoracsur.2004.09.040] [Medline: 15620907]

19. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part ii. Statistical methods of meta-analysis. Korean J Radiol 2015;16(6):1188-1196 [FREE Full text] [doi: 10.3348/kjr.2015.16.6.1188] [Medline: 26576107]

20. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. Clin Epidemiol Glob Health 2019;7(2):192-198. [doi: 10.1016/j.cegh.2018.05.005]

21. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 2004 May 15;23(9):1351-1375. [doi: 10.1002/sim.1761] [Medline: 15116347]

22. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008 Dec 16;149(12):889-897 [FREE Full text] [doi: 10.7326/0003-4819-149-12-200812160-00008] [Medline: 19075208]

23. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, IBIS Network, Clinical Sites, Data Coordinating Center, Image Processing Core, Statistical Analysis. Early brain development in infants at high risk for autism spectrum disorder. Nature 2017 Feb 15;542(7641):348-351 [FREE Full text] [doi: 10.1038/nature21369] [Medline: 28202961]

24. Chaddad A, Desrosiers C, Hassan L, Tanougast C. Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder. BMC Neurosci 2017 Jul 11;18(1):52 [FREE Full text] [doi: 10.1186/s12868-017-0373-0] [Medline: 28821235]

25. Wee C, Wang L, Shi F, Yap P, Shen D. Diagnosis of autism spectrum disorders using regional and interregional morphological features. Hum Brain Mapp 2014 Jul;35(7):3414-3430 [FREE Full text] [doi: 10.1002/hbm.22411] [Medline: 25050428]

26. Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, MRC AIMS Consortium. Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. Neuroimage 2010 Jan 1;49(1):44-56. [doi: 10.1016/j.neuroimage.2009.08.024] [Medline: 19683584]

27. Ecker C, Marquand A, Mourão-Miranda J, Johnston P, Daly EM, Brammer MJ, et al. Describing the brain in autism in five dimensions--magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. J Neurosci 2010 Aug 11;30(32):10612-10623 [FREE Full text] [doi: 10.1523/JNEUROSCI.5413-09.2010] [Medline: 20702694]

28. Xiao X, Fang H, Wu J, Xiao C, Xiao T, Qian L, et al. Diagnostic model generated by MRI-derived brain features in toddlers with autism spectrum disorder. Autism Res 2017 Apr;10(4):620-630. [doi: 10.1002/aur.1711] [Medline: 27874271]

29. Katuwal GJ, Cahill ND, Baum SA, Michael AM. The predictive power of structural MRI in autism diagnosis. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:4270-4273. [doi: 10.1109/EMBC.2015.7319338] [Medline: 26737238]

30. Jiao Y, Chen R, Ke X, Chu K, Lu Z, Herskovits EH. Predictive models of autism spectrum disorder based on brain regional cortical thickness. Neuroimage 2010 Apr 1;50(2):589-599 [FREE Full text] [doi: 10.1016/j.neuroimage.2009.12.047] [Medline: 20026220]

31. Neeley ES, Bigler ED, Krasny L, Ozonoff S, McMahon W, Lainhart JE. Quantitative temporal lobe differences: autism distinguished from controls using classification and regression tree analysis. Brain Dev 2007 Aug;29(7):389-399 [FREE Full text] [doi: 10.1016/j.braindev.2006.11.006] [Medline: 17204387]

32. Kong Y, Gao J, Xu Y, Pan Y, Wang J, Liu J. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. Neurocomputing 2019 Jan;324:63-68. [doi: 10.1016/j.neucom.2018.04.080]

33. Shen MD, Nordahl CW, Li DD, Lee A, Angkustsiri K, Emerson RW, et al. Extra-axial cerebrospinal fluid in high-risk and normal-risk children with autism aged 2-4 years: a case-control study. Lancet Psychiatry 2018 Nov;5(11):895-904. [doi: 10.1016/S2215-0366(18)30294-3] [Medline: 30270033]

34. Li H, Parikh NA, He L. A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. Front Neurosci 2018;12:491 [FREE Full text] [doi: 10.3389/fnins.2018.00491] [Medline: 30087587]

35. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. Neuroimage Clin 2018;17:16-23 [FREE Full text] [doi: 10.1016/j.nicl.2017.08.017] [Medline: 29034163]

36. Dekhil O, Hajjdiab H, Shalaby A, Ali MT, Ayinde B, Switala A, et al. Using resting state functional MRI to build a personalized autism diagnosis system. PLoS One 2018;13(10):e0206351 [FREE Full text] [doi: 10.1371/journal.pone.0206351] [Medline: 30379950]

37. Bernas A, Aldenkamp AP, Zinger S. Wavelet coherence-based classifier: a resting-state functional MRI study on neurodynamics in adolescents with high-functioning autism. Comput Methods Programs Biomed 2018 Feb;154:143-151 [FREE Full text] [doi: 10.1016/j.cmpb.2017.11.017] [Medline: 29249338]

38. Mastrovito D, Hanson C, Hanson SJ. Differences in atypical resting-state effective connectivity distinguish autism from schizophrenia. Neuroimage Clin 2018;18:367-376 [FREE Full text] [doi: 10.1016/j.nicl.2018.01.014] [Medline: 29487793]

39.   Emerson RW, Adams C, Nishino T, Hazlett HC, Wolff JJ, Zwaigenbaum L, IBIS Network, et al. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. Sci Transl Med 2017 Jun 7;9(393) [FREE Full text] [doi: 10.1126/scitranslmed.aag2882] [Medline: 28592562]

40.   Price T, Wee CY, Gao W, Shen D. Multiple-network classification of childhood autism using functional connectivity dynamics. Med Image Comput Comput Assist Interv 2014;17(Pt 3):177-184. [doi: 10.1007/978-3-319-10443-0_23] [Medline: 25320797]

41.   Uddin LQ, Supekar K, Lynch CJ, Khouzam A, Phillips J, Feinstein C, et al. Salience network-based classification and prediction of symptom severity in children with autism. JAMA Psychiatry 2013 Aug;70(8):869-879 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.104] [Medline: 23803651]

42.   Wang H, Chen C, Fushing H. Extracting multiscale pattern information of fMRI based functional brain connectivity with application on classification of autism spectrum disorders. PLoS One 2012;7(10):e45502 [FREE Full text] [doi: 10.1371/journal.pone.0045502] [Medline: 23056205]

43.   Iidaka T. Resting state functional magnetic resonance imaging and neural network classified autism and control. Cortex 2015 Feb;63:55-67. [doi: 10.1016/j.cortex.2014.08.011] [Medline: 25243989]

44.   Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316 [FREE Full text] [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]

45.   Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]

Abbreviations

ASD: autism spectrum disorder
AUC: area under the curve
DNN: deep neural network
DSM: Diagnostic and Statistical Manual of Mental Disorders
EEG: electroencephalogram
fMRI: functional magnetic resonance imaging
FN: false negative
FP: false positive
ICD: International Statistical Classification of Diseases
IEEE: Institute of Electrical and Electronics Engineers
MRI: magnetic resonance imaging
pAUC: partial AUC
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies-2
RoB: risk of bias
sMRI: structural magnetic resonance imaging
SROC: Summary Receiver Operating Characteristics
STARD: Standards for Reporting of Diagnostic Accuracy Studies
SVM: support vector machine
TN: true negative
TP: true positive

Review

How New Technologies Can Improve Prediction, Assessment, and Intervention in Obsessive-Compulsive Disorder (e-OCD): Review

Florian Ferreri[1*], MD, PhD; Alexis Bourla[1,2*], MD; Charles-Siegfried Peretti[1], MD, PhD; Tomoyuki Segawa[1], MD; Nemat Jaafari[3], MD, PhD; Stéphane Mouchabac[1], MD

[1]Sorbonne Université, Department of Adult Psychiatry and Medical Psychology, APHP, Saint-Antoine Hospital, Paris, France

[2]Jeanne d'Arc Hospital, INICEA Group, Saint Mandé, France

[3]INSERM, Pierre Deniker Clinical Research Unit, Henri Laborit Hospital & Experimental and Clinical Neuroscience Laboratory, Poitiers University Hospital, Poitier, France

[*]these authors contributed equally

Corresponding Author:
Alexis Bourla, MD
Sorbonne Université
Department of Adult Psychiatry and Medical Psychology
APHP, Saint-Antoine Hospital
184 rue du Faubourg Saint-Antoine
Paris, 75012
France
Phone: 33 149282635
Email: alexis.bourla@aphp.fr

Abstract

Background: New technologies are set to profoundly change the way we understand and manage psychiatric disorders, including obsessive-compulsive disorder (OCD). Developments in imaging and biomarkers, along with medical informatics, may well allow for better assessments and interventions in the future. Recent advances in the concept of digital phenotype, which involves using computerized measurement tools to capture the characteristics of a given psychiatric disorder, is one paradigmatic example.

Objective: The impact of new technologies on health professionals' practice in OCD care remains to be determined. Recent developments could disrupt not just their clinical practices, but also their beliefs, ethics, and representations, even going so far as to question their professional culture. This study aimed to conduct an extensive review of new technologies in OCD.

Methods: We conducted the review by looking for titles in the PubMed database up to December 2017 that contained the following terms: [Obsessive] AND [Smartphone] OR [phone] OR [Internet] OR [Device] OR [Wearable] OR [Mobile] OR [Machine learning] OR [Artificial] OR [Biofeedback] OR [Neurofeedback] OR [Momentary] OR [Computerized] OR [Heart rate variability] OR [actigraphy] OR [actimetry] OR [digital] OR [virtual reality] OR [Tele] OR [video].

Results: We analyzed 364 articles, of which 62 were included. Our review was divided into 3 parts: prediction, assessment (including diagnosis, screening, and monitoring), and intervention.

Conclusions: The review showed that the place of connected objects, machine learning, and remote monitoring has yet to be defined in OCD. Smartphone assessment apps and the Web Screening Questionnaire demonstrated good sensitivity and adequate specificity for detecting OCD symptoms when compared with a full-length structured clinical interview. The ecological momentary assessment procedure may also represent a worthy addition to the current suite of assessment tools. In the field of intervention, CBT supported by smartphone, internet, or computer may not be more effective than that delivered by a qualified practitioner, but it is easy to use, well accepted by patients, reproducible, and cost-effective. Finally, new technologies are enabling the development of new therapies, including biofeedback and virtual reality, which focus on the learning of coping skills. For them to be used, these tools must be properly explained and tailored to individual physician and patient profiles.

XSL·FO
RenderX

Introduction

Background

Obsessive-compulsive disorder (OCD) is a severe and frequent disorder with an estimated lifetime prevalence of 2.3% [1]. It has a poor outcome, with a remission rate of just 53% (95% CI 42-65) [2]. OCD typically runs a chronic course, with sequential periods of remission and relapse, and is associated with disabling comorbidities, including major depressive disorder (15%), social anxiety disorder (14%), generalized anxiety disorder (13%), persistent depressive disorder (13%), tic disorder (12.5%), body dysmorphic disorder (8.71%), and self-harming behavior (7.43%) [3]. Major functional and emotional impairments are often seen, with an impact on quality of life. In this context, being able to predict the outcome, accurately assess OCD, and intervene in OCD are a major health issue.

New technologies are set to profoundly change our way of practicing psychiatry. At the interface of e-health, new technologies, and clinical observation, a large number of new tools are currently being developed for the assessment and treatment of several psychiatric disorders. Applied to the OCD field, we can talk about "e-OCD" for "e-health technology applied to OCD" just as we can talk about e-PTSD (for posttraumatic stress disorder) or e-Addictology (in the field of addictive disorders). Clinicians currently rely on conventional assessment methods, based on the systematic collection of clinical data during consultations, sometimes using standardized assessment tools (eg, Yale-Brown Obsessive-Compulsive Scale, Y-BOCS) [4]. New tools are disrupting this *classical* psychiatry. The recent development of *digital phenotyping* [5], which involves the extraction of psychiatric disorder characteristics by computerized measurement tools via a smartphone or connected device, is one paradigmatic example. Several digital phenotype models are now emerging for schizophrenia [6] and mood disorders [7]. Some behaviors can now be digitally objectified: hyperactivity can be picked up by an accelerometer, manic graphorrhea (a symptom of motor excitement exhibited as continual and incoherent writing) manifests itself in an increase in the number of text messages sent, and even reckless spending can be monitored by a smartphone app. For example, the frequent or excessive hand washing that occurs in some people with OCD could be assessed with a connected wristband.

In OCD, detailed behavioral assessments have been made possible by the ubiquitous use of smartphones to collect large amounts of data that, until recently, were not available to psychiatrists. New methods of data collection can be classified as either active or passive [8]: *active* (or live) data collection refers to all self-assessment procedures that can be implemented on a computer or smartphone (requiring input from the patient); *passive* data collection, via motion detection, smartphone use, and so on, involves background tasks in which patients do not know when data are being collected, thus minimizing the observer's influence. In the field of care, the expansion of Web-based and smartphone-based interventions holds out the prospect of having a therapist in the pocket [9], and the accessibility of virtual reality (VR) [10] also appears useful in OCD. Machine learning (ML), a special form of artificial intelligence (AI) that classifies data according to a number of variables, allows of patterns to be identified that can then be used to predict treatment outcome [11].

Objective

Recent reviews of how new technologies can improve prediction, assessment, and intervention in posttraumatic stress disorder (PTSD) [12] or addictive behaviors [13] have shown that, although they have the ability to profoundly change the way we practice psychiatry, the likely impact of these new technologies on health professionals' practice has yet to be determined [14]. All these innovations, although being at very different stages of development, have the potential to disrupt current practices. To inform health care practitioners about the opportunities and future challenges offered by these new technologies, and their shortcomings, we conducted a review of the technologies that can be used for treatment outcome prediction, assessment, and intervention in OCD.
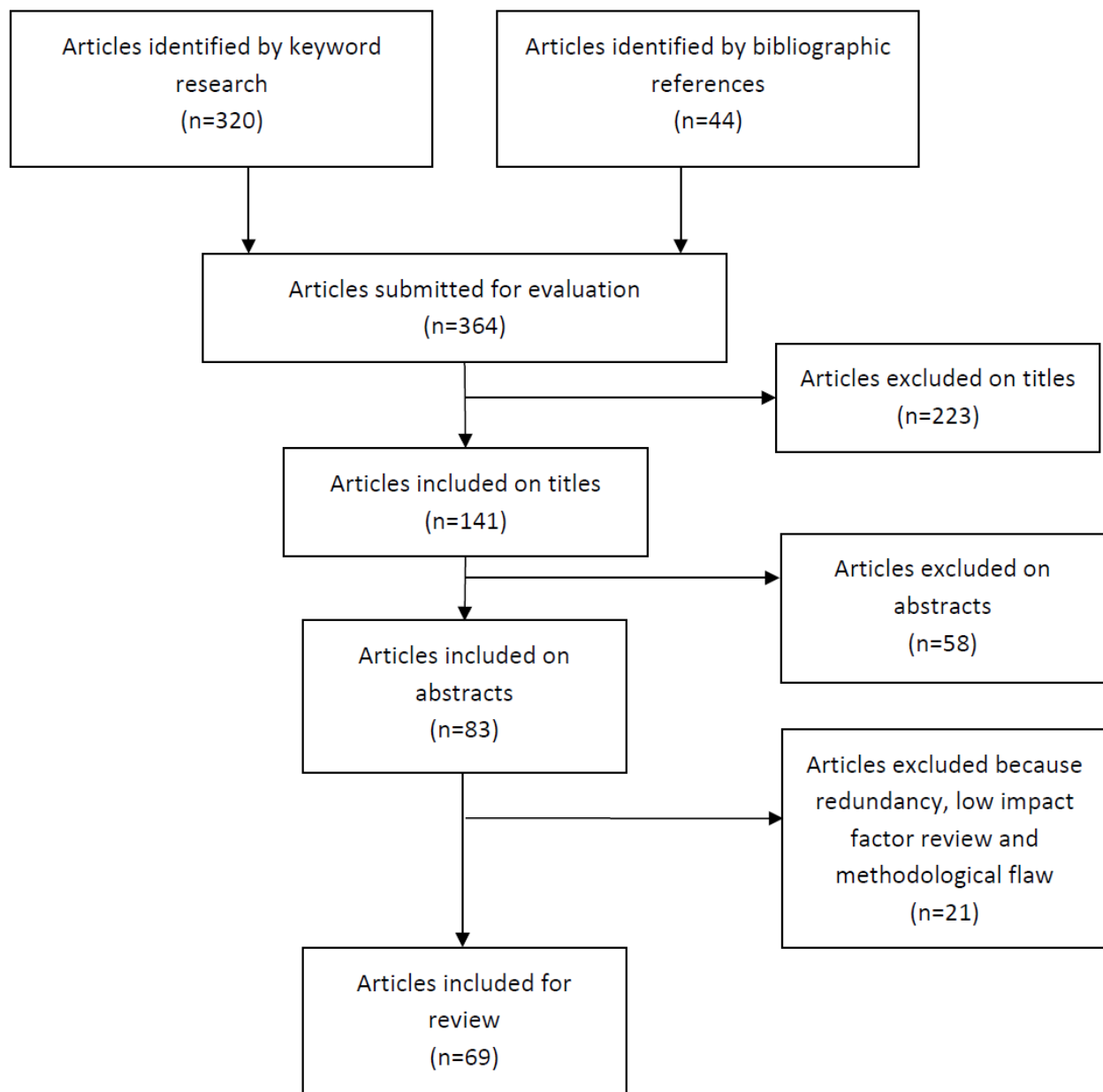
Methods

We conducted a review by looking for titles in the PubMed database up to December 2017 that contained the following terms: [Obsessive] AND [Smartphone] OR [phone] OR [Internet] OR [Device] OR [Wearable] OR [Mobile] OR [Machine learning] OR [Artificial] OR [Biofeedback] OR [Neurofeedback] OR [Momentary] OR [Computerized] OR [Heart rate variability] OR [actigraphy] OR [actimetry] OR [digital] OR [virtual reality] OR [Tele] OR [video].

The following inclusion criteria were used to identify studies involving OCD prediction, assessment, and intervention: eHealth apps (ie, computer-, smartphone-, or tablet-based apps), including telemedicine (ie, videoconferencing or phone-delivered cognitive behavioral therapy, CBT); wearable device (ie, smartphone sensor or electrocardiogram, ECG); or machine learning–based categorization.

Exclusion criteria were OCD symptoms not included as a primary or secondary outcome measure and focus on conventional media (ie, television, radio, or telephone).

Two authors, AB and FF, separately screened 364 abstracts, and 62 articles were included. This narrative review was divided into 3 parts: assessment, intervention, and treatment outcome prediction. See the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram in Figure 1.

This paper is not about a study that included patients. It was therefore not submitted to an ethics committee.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram.

Results

Given the fact that those technologies were at different stages of development, we present them from the most developed to the least developed.

Assessment

Three new assessment methods emerged from our review: electronic health (eHealth)-based assessment (using smartphone or internet apps), VR, and ML classification.

Electronic Health–Based Assessment

Smartphone assessment apps can either use a digital version of a validated scale or an ecological momentary assessment (EMA). Several apps are available or under development, but no study has so far validated assessment apps for OCD, and the use of Y-BOCS apps is not yet supported by research evidence [15].

The administration of psychological questionnaires via the internet is another form of assessment that may reduce the burden on patients and providers. These can be used for screening or for monitoring symptoms. The Web Screening Questionnaire is based on the Y-BOCS and can be completed in 2 min. It was found to have good sensitivity and adequate specificity for the detection of OCD symptoms when compared with a full-length structured clinical interview [16,17]. OCD can be identified using a Web-based test that mimics a structured interview: the Web-Based Depression and Anxiety Test has adequate sensitivity for OCD diagnosis (0.71) and good specificity (0.97), when compared with the Structured Clinical Interview for Diagnostic and Statistical Manual Mental Disorders [18]. Furthermore, the electronic Psychological Assessment System, a Web-based, self-report, multidisorder, clinical assessment and referral system showed a fair agreement with the clinical interview for OCD (kappa=0.39) [19]

With an EMA, patients respond to a series of questions based on their reported symptoms. This evaluation of symptoms daily in the patient's usual environment is theoretically free from recall biases, as the patient self-assesses "right then, not later; right there, not elsewhere" [20]. This new method using active data (supplied by the patient) profoundly modifies the assessment procedure, by introducing a computerized third party between a doctor and patient. The use of dedicated smartphone apps allows patients to keep an accurate diary of their symptoms and behaviors. In a clinical case study of EMA use in 3 adults diagnosed with OCD, Tilley and Rees [21] compared the numbers and types of obsessions and compulsions captured using the Y-BOCS, compared with an SMS-based EMA. Participants were told to record their experiences across a 12-hour period, in reply to text message prompts. The EMA approach yielded a lower number of obsessive and compulsive symptoms but produced additional types of obsessions and compulsions that had not been identified before by the Y-BOCS. The authors concluded that the EMA-OCD procedure may represent a worthy addition to the suite of assessment tools but requires research with larger samples.

Virtual Reality–Based Assessment

There are many definitions of VR, the most common being that it "refers to immersive, interactive, multi-sensory, viewer-centered, three-dimensional computer-generated environments and the combination of technologies required to build these environments" [22]. There is growing interest in VR as a key tool for investigating and assessing psychiatric disorders, as shown by the number of scientific articles with the term *virtual reality* in their title or abstract that are published in MEDLINE every year on this topic, which increased from 5 in 1991 to 842 in 2017.

Until recently, VR was limited by its cost and by the quality of the multimedia display technology. The recent democratization of these systems (Oculus Rift, PlayStation VR, HTC Vive, Samsung Gear) means that VR can now be used to perform neuropsychological assessments in real time [23]. The environment and the perceptual stimuli can be manipulated to trigger pathological behaviors (eg, checking behavior). This allows the clinician to assess behavioral responses to a situation that can elicit distress and to train patients how to cope with their symptoms.

Most research has focused on environmental trigger disorders (anxiety disorders in particular) [24], and we only found 3 studies of VR-based OCD assessment. This is difficult to understand, as this disorder seems an ideal candidate for this method, insofar as it is characterized by obsessions that are regularly induced by environmental triggers. A research group showed that the degree of VR-generated anxiety of individuals with OCD is positively correlated with their symptom and immersive tendency scores, suggesting that VR technology could be a useful anxiety-provoking tool [25]. When the same team replicated their study, they demonstrated that patients with OCD had significantly greater problems with compulsive checking than controls in a VR environment. Their results support the use of VR as a possible new behavioral measure of compulsive checking behavior [26].

Van Bennekom et al created a VR game from a first-person perspective that allowed patients to walk through a house where 15 OCD-related items were displayed, while simultaneously measuring their OCD symptoms (compulsions, anxiety, tension, uncertainty, and urge to control). A pilot study comparing 8 patients with OCD and 8 healthy controls showed that VR is capable of triggering OCD symptoms in patients, allowing clinicians to directly observe and assess OCD symptoms [27].

Machine Learning–Based Assessment

ML is the scientific discipline that focuses on how computers learn from data, using statistics to find relationships between them and efficient computing algorithms to accurately detect classification patterns [28]. This form of AI uses 2 different kinds of classification process: supervised and unsupervised. The former identifies rules from databases containing cases that have already been validated, while the latter looks for patterns in unlabeled data to find new structures. The coupling of ML with magnetic resonance imaging (MRI), electroencephalogram (EEG), or even blood tests can reveal patterns that allow patients to be divided into different groups (eg, patients at risk of relapse or patients with active disease). For example, ML is already showing considerable promise for predicting psychotic transition in patients in an at-risk mental state [29], and in the field of mood disorders [30].

We identified 5 studies in OCD, of which 3 were related to neuroanatomical data, 1 to EEG, and 1 to a set of both clinical and imaging data.

Hoexter et al applied support vector regression (SVR) to cortical volumes in individual structural MRI datasets of patients with OCD to predict symptom severity by identifying neurobiological markers. They found that the left medial orbitofrontal cortex (OFC) and left putamen contained the most discriminative information. Pearson correlation coefficient between predicted (with SVR) and observed (with Y-BOCS) symptom severity scores was 0.44 (P=.006), which is considered to be moderately positive [31].

Another study analyzed cortical and subcortical structures from MRI data of 38 patients with OCD and 36 controls and used different ML algorithms either without or with a feature selection algorithm (which is the process of selecting a subset of relevant features such as variables or predictors for use in model construction) to achieve an accurate distinction between patients and controls. Classification accuracy ranged from 52.56% (no feature selection) to 71.64% (t test with feature selection) [32].

Hu et al studied the application of multivariate pattern analysis to high-resolution T1-weighted MRI images acquired from 33 patients with OCD and 33 controls. The highest classification accuracy (81.82%) was achieved by a support vector machine (SVM) classifier using white matter information [33].

ML-based EEG classification is gaining interest for several psychiatric disorders, as it has very high classification accuracy among mood disorders, up to 98.95% classification accuracy for bipolar disorder versus schizophrenia [34], and 80.19% for unipolar versus bipolar depressive disorder [30]. High classification accuracy has also been found for OCD. Examining

single-channel EEG and 2-channel interhemispheric dependency measurements to distinguish between patients with OCD (n=10) and controls (n=10), using SVM classifiers, resulted in 85% (SD 5.2) classification accuracy [35].

Mas et al applied 2 supervised classification methods of ML-based class prediction to a dataset that included structural MRI, diffusion tensor imaging, neuropsychological, and genetic (single-nucleotide polymorphisms) data to predict early onset OCD severity. Their model classified child and adolescent patients with OCD by disease severity with an accuracy of 90% in the test set and 70% in the validation sample [36].

Interventions

Many treatments are offered in daily practice, but only 2 have been scientifically validated for OCD: pharmacotherapy (antidepressant) and CBT, including exposure and response prevention (ERP). It is easy to digitally transpose CBT using personalized modular programs in a smartphone app, on the internet, or in a VR program (for a review see Aboujaoude [37]). Technologies can also be used to augment traditional CBT [38,39], and could address issues of accessibility and effectiveness by increasing treatment adherence. Some authors refer to this type of treatment as *tCBT*, for technology-empowered CBT, or *cCBT*, for computerized CBT [40]. A review assessing the comparative efficacy of tCBT versus therapist-administered CBT (TA-CBT) found that tCBT and TA-CBT did not differ significantly in their efficacy on OCD symptoms, although there was a trend favoring TA-CBT [41]. Their study used mixed different technologies (self-help books, leaflets, bibliotherapy, internet, webcams, telephones, phone-delivered interactive voice response systems, and CD-ROMs), thereby making it impossible to disentangle the specific effects of the new technologies (smartphone in particular) delivering CBT. In addition, biofeedback, a new type of treatment using a computer, is starting to emerge, but was not included in any of the previous reviews [37].

Articles were divided in 5 categories: smartphone-based, Web-based (online program) or computer-based (ie, software) interventions, VR, and biofeedback. These categories were chosen because they correspond to the most innovative topics and are most often reported in studies of new technologies in psychiatry.

Telemedicine-Based Interventions

We found 9 studies, 2 case series reports, and 6 trials. All trials concluded that the videoconference intervention was effective and very acceptable for patients with OCD. However, face-to-face control design studies are lacking. One open-label trial concluded that a videoconferencing-based intervention is effective in the treatment of OCD in adults [42]. Among 5 randomized trials, 2 focused on the young with OCD [43,44], with 1 trial having control waitlist, and 1 trial having control consisting of face-to-face family-based CBT. The results are summarized in Table 1.

Table 1. Videoconferencing-based interventions.

Study, year	Aim	Methods	Results
Storch et al [43], 2011	To assess the efficacy of a 12-week webcam-delivered CBT[a] with therapist (14 sessions of 60-90 min) plus a brief initial face-to-face session in young OCDs[b].	31 youth with OCD (range=7-16 years; 19 male) were randomly assigned to webcam-delivered CBT or a waitlist control. Assessments were conducted immediately before and after treatment, and at a 3-month follow-up (for webcam-delivered CBT arm only). Primary outcomes included CY-BOCS[c], CGI[d] rates, and remission status.	Webcam-delivered CBT was superior to the waitlist control on all primary outcome measures with large effect sizes (Cohen $d \geq 1.36$) and at follow-up within group (Cohen $d \geq 1.98$). In all, 13 of 16 participants of the treatment group reached at least 30% CY-BOCS reduction and were considered as responders.
Goetter et al [42], 2014	To assess the efficacy of videoconference-mediated (16-18 twice-weekly, 90-min, individual sessions with between-session phone check-ins), in adult OCDs.	Open trial involving 15 participants. Assessment at four different times: pretreatment, midtreatment, posttreatment, and 3-month follow-up. Primary outcomes included Y-BOCS[e], CGI rate and Quality of Life Enjoyment and Satisfaction Questionnaire short form score.	The pre and post effect sizes were significant for the OCD symptom severity (Hedges g=2.56) and quality of life (Hedges g=1.27), and 80% of participants were rated as very much or much improved on the CGI. A total of 30% participants no longer met DSM-IV-TR[f] criteria for OCD among the 10 individuals who completed the 3-month follow-up assessment.
Vogel et al [45], 2014	To assess the efficacy and working alliances of technology along with telephone calls in adult OCDs: 6 tablet-based videoconferencing sessions (N=6) or studio-based videoconference (N=4), and 9 telephone sessions.	30 adults were randomized to 12-week video-conference-assisted ERP[g] (N=10), self-help ERP (N=10), or a waitlist condition (N=10). Primary outcome included Y-BOCS score, and Working Alliance Inventory. Assessments were conducted before and after treatment (12 weeks) by a psychologist blinded to treatment condition.	Videoconferencing treatment produced significantly greater reductions in obsessive-compulsive symptoms compared with the 2 control conditions (post hoc analysis videoconferencing treatment compared with self-help, P=.01, and waitlist, P=.01). Patients rated the videoconferencing format as natural and reported strong working alliances with their therapists.
Herbst et al [46], 2014	To assess the efficacy of a Web-based writing therapy with therapeutic interaction based on the concept of CBT (iCBT[h]; 8-week treatment of 14 sessions) for adult OCD.	34 adults were randomized according to a waitlist control design with follow-up measures at 8 weeks and 6 months. The primary outcome was the change in the severity of OCD symptoms (Y-BOCS, self-report, and Obsessive-Compulsive Inventory-Revised OCI-R).	iCBT treatment produced significantly greater reductions in obsessive-compulsive symptoms compared with the waitlist control group (Cohen d=0.82 Y-BOCS SR and d=0.87 OCI-R), using an intention-to-treat analysis. This effect remained stable at the 6-month follow-up. Of the 30 completers, 90% rated their condition as improved and would recommend the program to their friends.
Herbst et al [47], 2016	To assess the patient-therapist relation of a Web-based writing therapy with therapist guidance based on the concept of CBT (iCBT; 8-week treatment of 14 sessions) for adult OCD.	30 adults were randomized according to a waitlist control design with follow-up measures at 8 weeks. Primary outcome focused on Working Alliance Inventory self-report posttreatment within group.	The posttreatment Working Alliance Inventory-SR composite score represented 77% of the maximum scale value, which indicates a high working alliance.
Comer et al [44], 2017	To assess the working alliance and treatment satisfaction of videoconferencing in early OCD youth.	22 young patients with OCD (aged between 4-8 years) were randomized into videoconference-delivered family-based CBT or clinic-based family-based CBT. Pre-and posttreatment, and 6-month follow-up assessments masked to treatment condition. Primary outcomes were scores on working alliance and treatment satisfaction (Client Satisfaction Questionnaire).	Treatment alliance and satisfaction were high across conditions. At posttreatment, 72.7% of internet cases and 60% of clinic cases showed "excellent response," and at follow-up, 80% of internet cases and 66.7% of clinic cases showed "excellent response" (defined as a 1 or 2 on the CGI Scale). Differences between conditions on clinical significance responder status were nonsignificant.

[a]CBT: cognitive behavioral therapy.

[b]OCD: obsessive-compulsive disorder.

[c]CY-BOCS: Children's Yale-Brown Obsessive-Compulsive Scale.

[d]CGI: Clinical Global Impression Scale.

[e]Y-BOCS: Yale-Brown Obsessive-Compulsive Scale.

[f]DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders fourth edition.

[g]ERP: exposure and response prevention.

[h]iCBT: Web-based cognitive behavioral therapy.

Smartphone-Based Interventions

A recent review analyzed the current status of eHealth apps in the field of depressive and anxiety disorders, including OCD [15]. The authors highlighted the lack of a smartphone app in this field for intervention, as they only found one study—a case report that discussed the utility of a smartphone-based geofeedback app for patients with excessive outdoor checking behavior. This app delivers an audio message (recorded alarm signal) if a compulsive behavior is suspected (ie, if the patient takes too long to cover a predefined distance) [48]. The goal of reducing the time taken to get to appointments was reached (from 2 hours down to 20 min), not because of the app itself, but because the patient did not want to attract the attention of people around him.

In the wake of this review, an open pilot trial investigated the feasibility, acceptability, and efficacy of an ERP app (LiveOCDFree) among 21 patients with mild to moderate symptoms [49]. This showed that smartphone-guided ERP is feasible and acceptable, with high rates of retention and satisfaction, as participants reported significant improvements in OCD symptoms, showing 4.25-point improvement on self-report Y-BOCS ($F_{2,40}$=4.25, P=.02), and 3.96-point improvement on Beck Anxiety Inventory ($F_{1.31,26.13}$=3.96, P=.047), although not in depressive symptoms or quality of life. However, symptom amelioration only occurred within the first 6 weeks of this treatment, and there was no further improvement after midtreatment.

Web-Based Interventions

We found 19 studies on this topic. Most concluded that a Web-based treatment is effective, highly acceptable to patients with OCD [50], and may reduce barriers to treatment access [51]. The results are summarized in Multimedia Appendix 1 [52-62].

Regarding the patient-therapist relationship, a systematic review found no differences between Web-based CBT (iCBT) and face-to-face therapy [63].

Computer-Based Interventions

Few studies have specifically focused on computerized CBT for OCD, of which 6 [64-69] featured computer-driven telephone interview system (BT Steps). Y-BOCS effect sizes (mean Cohen d=0.84) were smaller than those for therapist-led CBT (mean Cohen d=1.22) [70]. The most recent study [68] randomly assigned 87 patients to 12 weeks of treatment with either BT Steps alone (n=28), BT Steps with nontherapist coaching (n=28), or BT Steps with CBT therapist coaching (n=31). All 3 interventions brought about a significant reduction in Y-BOCS scores, with effect sizes (Cohen d) of 1.16, 1.41, and 1.12, respectively. The main finding of this study was that when patients were asked which method of therapy (computer vs clinician) they preferred, they chose iCBT (computer: 48%; face-to-face therapy: 33%; and no stated preference: 19%). For a review see Lovell and Bee [71]. It should be noted that BT Steps was subsequently modified to be used online, and its name was changed to *OCFighter* [57], but Lovell et al found that, when used with low intensity, OCFighter does not lead to

clinically significant benefits, although it may reduce uptake of therapist-led CBT [72].

In another study, patients underwent three 45-min sessions at weekly intervals on an interactive computer program that provided vicarious exposure and response prevention for OCD, but no significant change in Y-BOCS scores was reported, although a significant change in depressive symptoms was observed [73].

One study focused on a computerized psychoeducative tool as an add-on to standard CBT, reporting variable acceptance across patients and no difference in the Y-BOCS score reduction when compared with a group using standard CBT alone [74].

Kalanthroff et al [75] developed a program called *Personalized-Computerized Inhibitory Training* (P-CIT) that sought to improve patients' ability to inhibit responses when exposed to images that were related to their specific OCD symptoms. They combined P-CIT and ERP in an 11-patient study involving training with P-CIT in 3 sessions of 15 min for 7 consecutive days, followed by 8 in-person 60-min ERP sessions with a trained therapist delivered over 2 or 3 weeks, all the while continuing P-CIT. The Y-BOCS score change was estimated over time, at weeks 0, 1, and 3. Y-BOCS scores decreased significantly over time (b=−3.47/week, t_{19}=−7.46, P<.001), and all patients save one achieved remission, with a Y-BOCS reduction of at least a 35% (mean reduction of 11 points).

Virtual Reality–Based Interventions

Kim et al [76] used a virtual environment to produce variations in arrangement anxiety in 24 patients with compulsive arranging symptoms. Patients performed virtual arrangement tasks 3 times, at 3-day intervals, and results showed that arrangement-related anxiety levels decreased significantly between the first and last days. No Y-BOCS scoring was used in this study.

Laforest et al [77] enrolled 3 adults with contamination OCD and exposed them to 2 virtual environments: a training environment (*neutral*) and an experimental (*contaminated*) environment. They assessed the presence and intensity of obsessions and compulsions (baseline, 3, 4, and 5 weeks, and at the end of a 12-session treatment). Exposure in VR (ie, touching walls and toilet bowls with varying degrees of filthiness) was discussed during a CBT session (reviewing the exposure session, performing cognitive restructuring of dysfunctional thoughts, and discussing upcoming homework assignments). The authors found a significant improvement in all 3 participants: pretreatment Y-BOCS scores were 22, 31, and 30, respectively, and at the 4-month follow-up were 16, 11, and 23, respectively. It should be noted that at the 8-month follow-up, Patient 2 still had a reduced Y-BOCS score, but the Y-BOCS scores of the other 2 patients had risen to 21 and 27.

Biofeedback-Based Interventions

Developed in the 1970s [78], biofeedback is a painless, noninvasive procedure that consists of capturing biometric data such as EEG, ECG, electromyogram (EMG), skin conductance, and temperature, and immediately feeding them back to the patient. The objective is to model the patient's brain activity in

XSL·FO

RenderX

real time as an image (video game type) or sound. Based on CBT techniques and relaxation, patients gradually learn to promote brain activity corresponding to the therapeutic target, through positive reinforcement. When activity in a desirable frequency band increases, the symbol modeling the brain activity changes in one direction, and when activity in an unfavorable band increases, the symbol changes in the opposite direction. Patients gradually learn the new brainwave, taking a wave corresponding to what is observed in healthy individuals as their model.

The first use of biofeedback in the field of OCD goes back to 1977, when a 25-year-old woman was treated with systematic desensitization in which EMG biofeedback was used to achieve relaxation [79].

Sürmeli and Ertem used quantitative EEG-guided neurofeedback (NF) in a case series of 36 treatment-resistant patients with OCD. The NF intervention consisted of inhibiting (keeping the activity below a set threshold) EEG theta or alpha rhythms on frontal, prefrontal and frontotemporal deviations. All participants underwent daily 60-min sessions for 9 to 84 days. Results showed that 91% of participants who received NF training showed a clinical improvement, according to the Y-BOCS and Clinical Global Impression (CGI) Scale, and 52% of participants maintained the improvements in their OCD symptoms at the 26-month follow-up [80].

Another study assessed NF efficacy in a randomized, double-blind, parallel design, involving 20 inpatients with OCD who underwent 25 sessions of either NF or sham (placebo) feedback (SF). The aim of the NF intervention was to reduce EEG activity in an independent component previously reported to be abnormal for this diagnosis. Although a pre- versus posttreatment comparison of the trained component and frequency did not yield significant results, the NF group had a significantly greater reduction in compulsions, compared with the SF group (P=.015) [81].

Deng et al used EEG biofeedback training as an adjunct to standardized treatment (antidepressant medication plus CBT) in a randomized controlled trial involving 79 patients with OCD. Of these, 40 were randomly assigned to the study group (antidepressant medication plus 8-week CBT plus NF sessions 5 times/week), and 39 were randomly assigned to the control group (antidepressant medication plus 8-week CBT). At 8 weeks, treatment was considered effective in 86.5% of participants in the study group and in 62.9% of participants in the control group, with mean decreases in the Y-BOCS score of 14.44 (study group) and 13.2 (control group)—a statistically significant (P=.003) but clinically irrelevant difference [82].

Machine Learning–Based Prediction

Five studies investigated whether ML approaches can predict treatment response or symptom severity.

Salomoni et al studied 130 participants under pharmacotherapy with selective serotonin reuptake inhibitors (SSRI, alone or SSRI plus low-dose risperidone) and/or CBT (ERP mostly), using 3 variables (Y-BOCS symptoms dimension, neuropsychological performances, and epidemiological data) to predict treatment outcome at 3 or 6 months. When a

multilayer perceptron (ie, supervised artificial neural network) was compared with a classic logistic regression model, it was found to have a considerably better predictive performance (93.3% vs 61.5%), when it came to correctly classifying patients as nonresponders to treatment (46.9% of participants) [83].

Another team used SVR to assess whether structural MRI assessing volumetric brain matter could predict symptom severity in 37 patients with no prior treatment. They found weak Pearson correlation coefficients (0.44-0.49) between observed and predicted severity using the Y-BOCS and dimensional Y-BOCS. The main value of this study was to highlight the ability of ML to identify neurobiological markers of OCD, as some regions contained more discriminative information than others: the left medial OFC) and left putamen were associated with severity, while the best predictors of the sexual/religious OCD subtype were the left medial OFC, right lateral OFC, and left anterior cingulate cortex [31].

Yun et al also used structural MRI to individualize biomarkers (cortical morphology) of treatment response to SSRI-based pharmacotherapy using SVM. A total of 56 treatment-naive patients with OCD and 75 healthy controls underwent T1-weighted MRI at baseline and after 4 months of newly introduced SSRI treatment (patients only). The SVM algorithm correctly classified the responders and nonresponders (based on variation in Y-BOCS score) with 90.7% to 95.6% accuracy (sensitivity=90.8%-96.2%; specificity=91.1%-95%) [84].

Askland et al used several variables (including Y-BOCS items, Neuroticism, Extraversion, and Openness to Experience-Five Factor Inventory items and subscale scores, Y-BOCS symptom checklist cleaning/washing compulsion score, and several self-report items from social adjustment scales), implemented in a Random Forest ML algorithm to predict remission outcome. Using 26 high-confidence features to predict a binary outcome (remitted vs never remitted), the algorithm correctly classified the patients (N=296) in 76.18% of cases (error rate=23.82%; bootstrap CI 22.10-25.45) [85].

Lenhard et al tested the ability of 4 different ML methods to predict treatment response to iCBT in a sample of 61 adolescents (12-17 years of age) with OCD. Participants were enrolled in a randomized controlled trial and received either immediate iCBT or delayed iCBT. The authors compared multivariate logistic regression with 4 ML algorithms (1 linear model with best subset predictor selection, and 3 flexible models: LASSO, Random Forest, and SVM) implemented with 46 demographic and clinical baseline variables (eg, Children's Y-BOCS score, OCD onset, OCD duration, symptom dimension, and CGI). The multivariate logistic regression was unable to detect significant predictors, whereas the 4 ML algorithms allowed treatment response to be predicted with between 75% (LASSO, Random Forest, and SVM) and 83% (linear model with best subset predictor selection) accuracy [86].

Discussion

Principal Findings

There is a constantly growing body of knowledge in the field of OCD care, and it is becoming increasingly complicated to

handle all these data on a daily basis. Our systematic review shows that the evaluation of new innovative technologies in the OCD is heterogeneous. Thus, the number of randomized studies available is low. With 5 randomized studies, telemedicine and the Web-based interventions are the most robustly evaluated.

In addition, more validated tools are needed to optimize the management of OCD. In the present review, we highlighted the diversity of new technologies used in psychiatry and their application to OCD for the purposes of prediction, diagnosis, and intervention. e-psychiatry is already booming, and some even talk about a digital mental health revolution [87]. These conclusions agree with the other reviews carried out in the field of psychiatric disorders (in particular PTSD and addictions) which highlight the interest of these new technologies [12,13].

Strengths and Limitations

However, EMA provide additional data but do not replace passing a scale with a trained clinician, and the reviewed studies does not bring systematic information about EMAs sensitivity and specificity. The timing (screening, evolution) or frequency of their use is not well defined. In addition, the absence of external contributor does not make it possible to ensure the veracity of the data collected.

As we have seen, VR seems very promising in OCD but in practice, in vivo exposure can hardly be proposed to all patients, and so does virtual exposure. Furthermore, an important limitation concerns the possibility of customizing the environment and the device tolerance since some side effects (such as dizziness, nausea, headache, and eyestrain) are not enough evaluated. As for the EMA, the optimal therapeutic protocol is not yet clearly defined (duration, number of sessions) as well as the duration in time of the effect, or the use in children. CBT supported by smartphone, internet, or computer clearly offers new therapeutic opportunities, but they may suffer from a lack of human interaction with more uncertain adherence (therapeutic alliance). The impossibility of finely adjusting coping strategies to progress is also a shortcoming. The question of self-administration of a treatment is also a significant risk. These risks seem lower with telemedicine, an older practice that also benefits from recent technological developments. It also provides solutions to the difficulties of access to care. Interaction with a health professional reduces the risk of poor compliance and allows better therapeutic adaptation.

In the end, few data are available on the severe forms of the disorder, classically excluded from studies.

The acceptability of these technologies must therefore be assessed at different levels [14,88]. This assessment is generally based on several major criteria: usability (device's flexibility and ease of learning), utility (technology's contribution), and satisfaction and reliability (including accuracy, effectiveness, and efficiency). Cost, though fundamental, is a secondary consideration. Finally, the concept of risk impinges on acceptability and constitutes an important dimension of medical reasoning. It must therefore be taken into account when these technologies are being assessed (impact of false positives or false negatives, ethical issues).

In addition to the concerns of acceptability, validity, cost, and data security, it is necessary to consider the specific problems of patients with OCD. The major component of this disorder is the verification ritual, which is performed to relieve the anxiety associated with obsessive ideas and intrusive doubt [89]. Even if this is not reported in studies, care should be taken to ensure that users do not develop digital stress or become a slave to an app or VR software.

The data on detection may not have the same impact as in other areas of psychiatry, in which delayed diagnoses are more common (eg, mood disorder or psychotic transition). However, new forms of assessment (EMA, VR- or ML-based assessment) perform better than standardized tests, and sometimes even better than therapist interviews.

The data that are currently available indicate that these new technologies could be extremely valuable in the field of treatment. Although they are not necessarily more effective than qualified practitioners, these new tools allow for the democratization of access to recommended therapies, including CBT and iCBT (via smartphone, internet, or computer), are easy to use, well accepted by patients, reproducible, and cost-effective. Enhanced CBT with VR makes exposure therapy possible without moving out of the clinical setting but needs a high degree of personalization to be effective. Furthermore, the VR could trigger the stimulus-response OCD pattern of the patient, and therefore, if not enough guided by a trained therapist, it could be a risk needed to be considered. NF, which is in an early stage development compared with other interventions, has however proved to be useful in reducing compulsion in some patients. Finally, in the field of prediction, severity or treatment success can be predicted with a good degree of accuracy. Being able to predict the response to a treatment will promote personalized medicine. Furthermore, given the current development of new technology in the psychiatric field, it seems important to increase training measures on these technologies, particularly by integrating them into the resident teaching program. Many ethical issues, data security, data storage, privacy, and hacking risk have yet to be resolved. Disease detection or risk prediction of OCD could be stressful for patients and brings the risk of excessive focus and anxious counter reaction. It is essential for psychiatrists to be involved in the development of these technologies, and developers have a major interest in communicating better about the design of these tools and the algorithms they want us to use in the near future.

Conclusions

Confidence in eHealth among patients with addictions and health care professionals is a major issue [90]. Studies have highlighted good acceptability and patient compliance. ML is revolutionizing fundamental research, by allowing for better classification of patients, based not only on clinical data but also on biological or neuroimaging-derived data. It is becoming reasonable to talk about genuinely complementary examinations in behavioral studies. Finally, these new technologies are enabling the development of new therapies, including biofeedback and VR, that focus on the learning of coping skills.

Authors' Contributions

AB and SM designed the study. AB and FF screened the abstracts and wrote the manuscript. NJ and CP helped to supervise the project and contributed to the editing. SM conceived the original idea and was in charge of overall direction and planning. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based intervention studies.
[DOCX File, 20 KB - mental_v6i12e11643_app1.docx]

References

1. Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry 2010 Jan;15(1):53-63 [FREE Full text] [doi: 10.1038/mp.2008.94] [Medline: 18725912]
2. Lochner C, Fineberg NA, Zohar J, van Ameringen M, Juven-Wetzler A, Altamura AC, et al. Comorbidity in obsessive-compulsive disorder (OCD): a report from the International College of Obsessive-Compulsive Spectrum Disorders (ICOCS). Compr Psychiatry 2014 Oct;55(7):1513-1519. [doi: 10.1016/j.comppsych.2014.05.020] [Medline: 25011690]
3. Sharma E, Thennarasu K, Reddy YC. Long-term outcome of obsessive-compulsive disorder in adults: a meta-analysis. J Clin Psychiatry 2014 Sep;75(9):1019-1027. [doi: 10.4088/JCP.13r08849] [Medline: 25295427]
4. Moritz S, Meier B, Kloss M, Jacobsen D, Wein C, Fricke S, et al. Dimensional structure of the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS). Psychiatry Res 2002 Mar 15;109(2):193-199. [doi: 10.1016/s0165-1781(02)00012-4] [Medline: 11927144]
5. Torous J, Onnela J, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 7;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
6. Torous J, Keshavan M. A new window into psychosis: the rise digital phenotyping, smartphone assessment, and mobile monitoring. Schizophr Res 2018 Jul;197:67-68. [doi: 10.1016/j.schres.2018.01.005] [Medline: 29338959]
7. Bourla A, Ferreri F, Ogorzelec L, Guinchard C, Mouchabac S. [Assessment of mood disorders by passive data gathering: the concept of digital phenotype versus psychiatrist's professional culture]. Encephale 2018 Apr;44(2):168-175. [doi: 10.1016/j.encep.2017.07.007] [Medline: 29096909]
8. 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]
9. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
10. Eichenberg C, Wolters C. IntechOpen. 2012. Virtual Realities in the Treatment of Mental Disorders: A Review of the Current State of Research. URL: https://www.intechopen.com/books/virtual-reality-in-psychological-medical-and-pedagogical-applications/virtual-realities-in-the-treatment-of-mental-disorders-a-review-of-the-current-state-of-research [accessed 2017-12-29]
11. Mouchabac S, Guinchard C. Contribution of machine learning in the prediction of transition to psychosis: what are the issues for the patient and the psychiatrist? L'Information psychiatrique 2013;89:811-817.
12. Bourla A, Mouchabac S, El Hage W, Ferreri F. e-PTSD: an overview on how new technologies can improve prediction and assessment of Posttraumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2018;9(sup1):1424448 [FREE Full text] [doi: 10.1080/20008198.2018.1424448] [Medline: 29441154]
13. Ferreri F, Bourla A, Mouchabac S, Karila L. e-Addictology: an overview of new technologies for assessing and intervening in addictive behaviors. Front Psychiatry 2018;9:51 [FREE Full text] [doi: 10.3389/fpsyt.2018.00051] [Medline: 29545756]
14. Bourla A, Ferreri F, Ogorzelec L, Peretti C, Guinchard C, Mouchabac S. Psychiatrists' attitudes toward disruptive new technologies: mixed-methods study. JMIR Ment Health 2018 Dec 14;5(4):e10240 [FREE Full text] [doi: 10.2196/10240] [Medline: 30552086]
15. van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]
16. Donker T, van Straten A, Marks I, Cuijpers P. A brief web-based screening questionnaire for common mental disorders: development and validation. J Med Internet Res 2009 Jul 24;11(3):e19 [FREE Full text] [doi: 10.2196/jmir.1134] [Medline: 19632977]

17. Coles ME, Cook LM, Blake TR. Assessing obsessive compulsive symptoms and cognitions on the internet: evidence for the comparability of paper and Internet administration. Behav Res Ther 2007 Sep;45(9):2232-2240. [doi: 10.1016/j.brat.2006.12.009] [Medline: 17306222]

18. Farvolden P, McBride C, Bagby R, Ravitz P. A web-based screening instrument for depression and anxiety disorders in primary care. J Med Internet Res 2003;5(3):e23 [FREE Full text] [doi: 10.2196/jmir.5.3.e23] [Medline: 14517114]

19. Nguyen DP, Klein B, Meyer D, Austin DW, Abbott JM. The diagnostic validity and reliability of an internet-based clinical assessment program for mental disorders. J Med Internet Res 2015 Sep 21;17(9):e218 [FREE Full text] [doi: 10.2196/jmir.4195] [Medline: 26392066]

20. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling method. J Nerv Ment Dis 1987 Sep;175(9):526-536. [doi: 10.1097/00005053-198709000-00004] [Medline: 3655778]

21. Tilley PJ, Rees CS. A clinical case study of the use of ecological momentary assessment in obsessive compulsive disorder. Front Psychol 2014;5:339 [FREE Full text] [doi: 10.3389/fpsyg.2014.00339] [Medline: 24860521]

22. Srivastava K, Das RC, Chaudhury S. Virtual reality applications in mental health: challenges and perspectives. Ind Psychiatry J 2014;23(2):83-85 [FREE Full text] [doi: 10.4103/0972-6748.151666] [Medline: 25788795]

23. Negut A, Matu S, Sava FA, David D. Virtual reality measures in neuropsychological assessment: a meta-analytic review. Clin Neuropsychol 2016 Feb;30(2):165-184. [doi: 10.1080/13854046.2016.1144793] [Medline: 26923937]

24. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]

25. Kim K, Kim C, Cha KR, Park J, Han K, Kim YK, et al. Anxiety provocation and measurement using virtual reality in patients with obsessive-compulsive disorder. Cyberpsychol Behav 2008 Dec;11(6):637-641. [doi: 10.1089/cpb.2008.0003] [Medline: 18991527]

26. Kim K, Kim SI, Cha KR, Park J, Rosenthal MZ, Kim J, et al. Development of a computer-based behavioral assessment of checking behavior in obsessive-compulsive disorder. Compr Psychiatry 2010;51(1):86-93. [doi: 10.1016/j.comppsych.2008.12.001] [Medline: 19932831]

27. van Bennekom MJ, Kasanmoentalib MS, de Koning PP, Denys D. A virtual reality game to assess obsessive-compulsive disorder. Cyberpsychol Behav Soc Netw 2017 Nov;20(11):718-722. [doi: 10.1089/cyber.2017.0107] [Medline: 29125791]

28. Deo RC. Machine learning in medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]

29. Koutsouleris N, Borgwardt S, Meisenzahl EM, Bottlender R, Möller HJ, Riecher-Rössler A. Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: results from the FePsy study. Schizophr Bull 2012 Nov;38(6):1234-1246 [FREE Full text] [doi: 10.1093/schbul/sbr145] [Medline: 22080496]

30. Erguzel T, Tas C, Cebi M. A wrapper-based approach for feature selection and classification of major depressive disorder-bipolar disorders. Comput Biol Med 2015 Sep;64:127-137. [doi: 10.1016/j.compbiomed.2015.06.021] [Medline: 26164033]

31. Hoexter MQ, Miguel EC, Diniz JB, Shavitt RG, Busatto GF, Sato JR. Predicting obsessive-compulsive disorder severity combining neuroimaging and machine learning methods. J Affect Disord 2013 Sep 25;150(3):1213-1216. [doi: 10.1016/j.jad.2013.05.041] [Medline: 23769292]

32. Trambaiolli L, Biazoli CJ, Balardin J, Hoexter M, Sato J. The relevance of feature selection methods to the classification of obsessive-compulsive disorder based on volumetric measures. J Affect Disord 2017 Nov;222:49-56. [doi: 10.1016/j.jad.2017.06.061] [Medline: 28672179]

33. Hu X, Liu Q, Li B, Tang W, Sun H, Li F, et al. Multivariate pattern analysis of obsessive-compulsive disorder using structural neuroanatomy. Eur Neuropsychopharmacol 2016 Feb;26(2):246-254. [doi: 10.1016/j.euroneuro.2015.12.014] [Medline: 26708318]

34. AliMardani F, Boostani R, Blankertz B. Presenting a Spatial-Geometric EEG feature to classify BMD and schizophrenic patients. Int J Adv Telecommun Electrotech Sig Syst 2016;5(2):79-85. [doi: 10.11601/ijates.v5i2.143]

35. Aydin S, Arica N, Ergul E, Tan O. Classification of obsessive compulsive disorder by EEG complexity and hemispheric dependency measurements. Int J Neural Syst 2015 May;25(3):1550010. [doi: 10.1142/S0129065715500100] [Medline: 25804351]

36. Mas S, Gassó P, Morer A, Calvo A, Bargalló N, Lafuente A, et al. Integrating genetic, neuropsychological and neuroimaging data to model early-onset obsessive compulsive disorder severity. PLoS One 2016;11(4):e0153846 [FREE Full text] [doi: 10.1371/journal.pone.0153846] [Medline: 27093171]

37. Aboujaoude E. Three decades of telemedicine in obsessive-compulsive disorder: a review across platforms. J Obsess-Compuls Rel 2017 Jul;14:65-70. [doi: 10.1016/j.jocrd.2017.06.003]

38. Lind C, Boschen M, Morrissey S. Technological advances in psychotherapy: implications for the assessment and treatment of obsessive compulsive disorder. J Anxiety Disord 2013 Jan;27(1):47-55. [doi: 10.1016/j.janxdis.2012.09.004] [Medline: 23247201]

39.   Kyrios M, Ahern C, Fassnacht DB, Nedeljkovic M, Moulding R, Meyer D. Therapist-assisted internet-based cognitive behavioral therapy versus progressive relaxation in obsessive-compulsive disorder: randomized controlled trial. J Med Internet Res 2018 Aug 8;20(8):e242 [FREE Full text] [doi: 10.2196/jmir.9566] [Medline: 30089607]

40.   Herbst N, Voderholzer U, Stelzer N, Knaevelsrud C, Hertenstein E, Schlegl S, et al. The potential of telemental health applications for obsessive-compulsive disorder. Clin Psychol Rev 2012 Aug;32(6):454-466. [doi: 10.1016/j.cpr.2012.04.005] [Medline: 22705583]

41.   Dèttore D, Pozza A, Andersson G. Efficacy of technology-delivered cognitive behavioural therapy for OCD versus control conditions, and in comparison with therapist-administered CBT: meta-analysis of randomized controlled trials. Cogn Behav Ther 2015;44(3):190-211. [doi: 10.1080/16506073.2015.1005660] [Medline: 25705787]

42.   Goetter EM, Herbert JD, Forman EM, Yuen EK, Thomas J. An open trial of videoconference-mediated exposure and ritual prevention for obsessive-compulsive disorder. J Anxiety Disord 2014 Jun;28(5):460-462. [doi: 10.1016/j.janxdis.2014.05.004] [Medline: 24873883]

43.   Storch EA, Caporino NE, Morgan JR, Lewin AB, Rojas A, Brauer L, et al. Preliminary investigation of web-camera delivered cognitive-behavioral therapy for youth with obsessive-compulsive disorder. Psychiatry Res 2011 Oct 30;189(3):407-412. [doi: 10.1016/j.psychres.2011.05.047] [Medline: 21684018]

44.   Comer JS, Furr JM, Cooper-Vince CE, Kerns CE, Chan PT, Edson AL, et al. Internet-delivered, family-based treatment for early-onset OCD: a preliminary case series. J Clin Child Adolesc Psychol 2014;43(1):74-87 [FREE Full text] [doi: 10.1080/15374416.2013.855127] [Medline: 24295036]

45.   Vogel PA, Solem S, Hagen K, Moen EM, Launes G, Håland AT, et al. A pilot randomized controlled trial of videoconference-assisted treatment for obsessive-compulsive disorder. Behav Res Ther 2014 Dec;63:162-168. [doi: 10.1016/j.brat.2014.10.007] [Medline: 25461792]

46.   Herbst N, Voderholzer U, Thiel N, Schaub R, Knaevelsrud C, Stracke S, et al. No talking, just writing! Efficacy of an internet-based cognitive behavioral therapy with exposure and response prevention in obsessive compulsive disorder. Psychother Psychosom 2014;83(3):165-175. [doi: 10.1159/000357570] [Medline: 24732962]

47.   Herbst N, Franzen G, Voderholzer U, Thiel N, Knaevelsrud C, Hertenstein E, et al. Working alliance in internet-based cognitive-behavioral therapy for obsessive-compulsive disorder. Psychother Psychosom 2016;85(2):117-118. [doi: 10.1159/000441282] [Medline: 26807976]

48.   Olbrich H, Stengler K, Olbrich S. Smartphone based Geo-Feedback in obsessive compulsive disorder as facilitatory intervention: a case report. J Obsess-Compuls Rel 2016 Jan;8:75-78. [doi: 10.1016/j.jocrd.2016.01.001]

49.   Boisseau C, Schwartzman C, Lawton J, Mancebo M. App-guided exposure and response prevention for obsessive compulsive disorder: an open pilot trial. Cogn Behav Ther 2017 Nov;46(6):447-458. [doi: 10.1080/16506073.2017.1321683] [Medline: 28565937]

50.   Lenhard F, Vigerland S, Engberg H, Hallberg A, Thermaenius H, Serlachius E. 'On My Own, but Not Alone' - Adolescents' experiences of internet-delivered cognitive behavior therapy for obsessive-compulsive disorder. PLoS One 2016;11(10):e0164311 [FREE Full text] [doi: 10.1371/journal.pone.0164311] [Medline: 27711249]

51.   Wootton BM, Titov N, Dear BF, Spence J, Kemp A. The acceptability of Internet-based treatment and characteristics of an adult sample with obsessive compulsive disorder: an Internet survey. PLoS One 2011;6(6):e20548 [FREE Full text] [doi: 10.1371/journal.pone.0020548] [Medline: 21673987]

52.   Andersson E, Ljótsson B, Hedman E, Kaldo V, Paxling B, Andersson G, et al. Internet-based cognitive behavior therapy for obsessive compulsive disorder: a pilot study. BMC Psychiatry 2011 Aug 3;11:125 [FREE Full text] [doi: 10.1186/1471-244X-11-125] [Medline: 21812991]

53.   Wootton B, Titov N, Dear B, Spence J, Andrews G, Johnston L, et al. An internet administered treatment program for obsessive-compulsive disorder: a feasibility study. J Anxiety Disord 2011 Dec;25(8):1102-1107. [doi: 10.1016/j.janxdis.2011.07.009] [Medline: 21899983]

54.   Andersson E, Enander J, Andrén P, Hedman E, Ljótsson B, Hursti T, et al. Internet-based cognitive behaviour therapy for obsessive-compulsive disorder: a randomized controlled trial. Psychol Med 2012 Oct;42(10):2193-2203 [FREE Full text] [doi: 10.1017/S0033291712000244] [Medline: 22348650]

55.   Lenhard F, Vigerland S, Andersson E, Rück C, Mataix-Cols D, Thulin U, et al. Internet-delivered cognitive behavior therapy for adolescents with obsessive-compulsive disorder: an open trial. PLoS One 2014;9(6):e100773 [FREE Full text] [doi: 10.1371/journal.pone.0100773] [Medline: 24949622]

56.   Andersson E, Steneby S, Karlsson K, Ljótsson B, Hedman E, Enander J, et al. Long-term efficacy of internet-based cognitive behavior therapy for obsessive-compulsive disorder with or without booster: a randomized controlled trial. Psychol Med 2014 Oct;44(13):2877-2887. [doi: 10.1017/S0033291714000543] [Medline: 25066102]

57.   Mahoney A, Mackenzie A, Williams A, Smith J, Andrews G. Internet cognitive behavioural treatment for obsessive compulsive disorder: a randomised controlled trial. Behav Res Ther 2014 Dec;63:99-106. [doi: 10.1016/j.brat.2014.09.012] [Medline: 25461784]

58.   Diefenbach G, Wootton B, Bragdon L, Moshier S, Tolin D. Treatment outcome and predictors of internet guided self-help for obsessive-compulsive disorder. Behav Ther 2015 Nov;46(6):764-774. [doi: 10.1016/j.beth.2015.06.001] [Medline: 26520219]

XSL•FO

RenderX

59. Rees CS, Anderson RA, Kane RT, Finlay-Jones AL. Online obsessive-compulsive disorder treatment: preliminary results of the 'OCD? Not Me!' self-guided internet-based cognitive behavioral therapy program for young people. JMIR Ment Health 2016 Jul 5;3(3):e29 [FREE Full text] [doi: 10.2196/mental.5363] [Medline: 27381977]

60. Seol S, Kwon JS, Kim YY, Kim SN, Shin M. Internet-based cognitive behavioral therapy for obsessive-compulsive disorder in Korea. Psychiatry Investig 2016 Jul;13(4):373-382 [FREE Full text] [doi: 10.4306/pi.2016.13.4.373] [Medline: 27482237]

61. Lenhard F, Andersson E, Mataix-Cols D, Rück C, Vigerland S, Högström J, et al. Therapist-guided, internet-delivered cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017 Jan;56(1):10-19.e2 [FREE Full text] [doi: 10.1016/j.jaac.2016.09.515] [Medline: 27993223]

62. Lenhard F, Ssegonja R, Andersson E, Feldman I, Rück C, Mataix-Cols D, et al. Cost-effectiveness of therapist-guided internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: results from a randomised controlled trial. BMJ Open 2017 May 17;7(5):e015246 [FREE Full text] [doi: 10.1136/bmjopen-2016-015246] [Medline: 28515196]

63. Sucala M, Schnur J, Constantino M, Miller S, Brackman E, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 2;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]

64. Greist J, Marks I, Baer L, Parkin J, Manzo P, Mantle J, et al. Self-treatment for obsessive compulsive disorder using a manual and a computerized telephone interview: a U.S.-U.K. study. MD Comput 1998;15(3):149-157. [Medline: 9617085]

65. Bachofen M, Nakagawa A, Marks I, Park J, Greist J, Baer L, et al. Home self-assessment and self-treatment of obsessive-compulsive disorder using a manual and a computer-conducted telephone interview: replication of a UK-US study. J Clin Psychiatry 1999 Aug;60(8):545-549. [doi: 10.4088/jcp.v60n0807] [Medline: 10485637]

66. Nakagawa A, Marks I, Park J, Bachofen M, Baer L, Dottl S, et al. Self-treatment of obsessive-compulsive disorder guided by manual and computer-conducted telephone interview. J Telemed Telecare 2000;6(1):22-26. [doi: 10.1258/1357633001933899] [Medline: 10824386]

67. Greist J, Marks I, Baer L, Kobak K, Wenzel K, Hirsch M, et al. Behavior therapy for obsessive-compulsive disorder guided by a computer or by a clinician compared with relaxation as a control. J Clin Psychiatry 2002 Feb;63(2):138-145. [doi: 10.4088/jcp.v63n0209] [Medline: 11874215]

68. Kenwright M, Marks I, Graham C, Franses A, Mataix-Cols D. Brief scheduled phone support from a clinician to enhance computer-aided self-help for obsessive-compulsive disorder: randomized controlled trial. J Clin Psychol 2005 Dec;61(12):1499-1508. [doi: 10.1002/jclp.20204] [Medline: 16173084]

69. Kobak K, Greist R, Jacobi D, Levy-Mack H, Greist J. Computer-assisted cognitive behavior therapy for obsessive-compulsive disorder: a randomized trial on the impact of lay vs. professional coaching. Ann Gen Psychiatry 2015;14:10 [FREE Full text] [doi: 10.1186/s12991-015-0048-0] [Medline: 25722737]

70. Tumur I, Kaltenthaler E, Ferriter M, Beverley C, Parry G. Computerised cognitive behaviour therapy for obsessive-compulsive disorder: a systematic review. Psychother Psychosom 2007;76(4):196-202. [doi: 10.1159/000101497] [Medline: 17570957]

71. Lovell K, Bee P. Optimising treatment resources for OCD: a review of the evidence base for technology-enhanced delivery. J Ment Health 2011 Dec;20(6):525-542. [doi: 10.3109/09638237.2011.608745] [Medline: 22126631]

72. Lovell K, Bower P, Gellatly J, Byford S, Bee P, McMillan D, et al. Low-intensity cognitive-behaviour therapy interventions for obsessive-compulsive disorder compared to waiting list for therapist-led cognitive-behaviour therapy: 3-arm randomised controlled trial of clinical effectiveness. PLoS Med 2017 Jun;14(6):e1002337 [FREE Full text] [doi: 10.1371/journal.pmed.1002337] [Medline: 28654682]

73. Kirkby K, Berrios G, Daniels B, Menzies R, Clark A, Romano A. Process-outcome analysis in computer-aided treatment of obsessive-compulsive disorder. Compr Psychiatry 2000;41(4):259-265. [doi: 10.1053/comp.2000.7431] [Medline: 10929793]

74. Morgiève M, N'Diaye K, Clair A, Pelissolo A, Mallet L. [Can the efficacy of behavioral and cognitive therapy for obsessive compulsive disorder be augmented by innovative computerized adjuvant?]. Encephale 2016 Oct;42(5):402-409. [doi: 10.1016/j.encep.2016.03.004] [Medline: 27109326]

75. Kalanthroff E, Steinman S, Schmidt A, Campeas R, Simpson H. Piloting a personalized computerized inhibitory training program for individuals with obsessive-compulsive disorder. Psychother Psychosom 2018;87(1):52-54. [doi: 10.1159/000481199] [Medline: 29306946]

76. Kim K, Roh D, Kim S, Kim C. Provoked arrangement symptoms in obsessive-compulsive disorder using a virtual environment: a preliminary report. Comput Biol Med 2012 Apr;42(4):422-427. [doi: 10.1016/j.compbiomed.2011.12.010] [Medline: 22226644]

77. Laforest M, Bouchard S, Bossé J, Mesly O. Effectiveness of in virtuo exposure and response prevention treatment using cognitive-behavioral therapy for obsessive-compulsive disorder: a study based on a single-case study protocol. Front Psychiatry 2016;7:99 [FREE Full text] [doi: 10.3389/fpsyt.2016.00099] [Medline: 27378951]

78. Kamiya J. Operant control of the EEG alpha rhythm and some of its reported effects on consciousness. In: Miller NE, Hnatiow M, Engel BT, editors. Biofeedback and Self-control: An Aldine Reader on the Regulation of Bodily Processes and Consciousness. Chicago, IL: Aldine-Atherton; 1971.

79. Delk JL. Use of EMG biofeedback in behavioral treatment of an obsessive-phobic-depressive syndrome. Dis Nerv Syst 1977 Nov;38(11):938-939. [Medline: 21059]

80. Sürmeli T, Ertem A. Obsessive compulsive disorder and the efficacy of qEEG-guided neurofeedback treatment: a case series. Clin EEG Neurosci 2011 Jul;42(3):195-201. [doi: 10.1177/155005941104200310] [Medline: 21870473]

81. Kopřivová J, Congedo M, Raszka M, Praško J, Brunovský M, Horáček J. Prediction of treatment response and the effect of independent component neurofeedback in obsessive-compulsive disorder: a randomized, sham-controlled, double-blind study. Neuropsychobiology 2013;67(4):210-223. [doi: 10.1159/000347087] [Medline: 23635906]

82. Deng X, Wang G, Zhou L, Zhang X, Yang M, Han G, et al. Randomized controlled trial of adjunctive EEG-biofeedback treatment of obsessive-compulsive disorder. Shanghai Arch Psychiatry 2014 Oct;26(5):272-279 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.214067] [Medline: 25477720]

83. Salomoni G, Grassi M, Mosini P, Riva P, Cavedini P, Bellodi L. Artificial neural network model for the prediction of obsessive-compulsive disorder treatment response. J Clin Psychopharmacol 2009 Aug;29(4):343-349. [doi: 10.1097/JCP.0b013e3181aba68f] [Medline: 19593173]

84. Yun J, Jang JH, Kim SN, Jung WH, Kwon JS. Neural correlates of response to pharmacotherapy in obsessive-compulsive disorder: individualized cortical morphology-based structural covariance. Prog Neuropsychopharmacol Biol Psychiatry 2015 Dec 3;63:126-133. [doi: 10.1016/j.pnpbp.2015.06.009] [Medline: 26116795]

85. Askland KD, Garnaat S, Sibrava NJ, Boisseau CL, Strong D, Mancebo M, et al. Prediction of remission in obsessive compulsive disorder using a novel machine learning strategy. Int J Methods Psychiatr Res 2015 Jun;24(2):156-169 [FREE Full text] [doi: 10.1002/mpr.1463] [Medline: 25994109]

86. Lenhard F, Sauer S, Andersson E, Månsson KN, Mataix-Cols D, Rück C, et al. Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. Int J Methods Psychiatr Res 2018 Mar;27(1). [doi: 10.1002/mpr.1576] [Medline: 28752937]

87. Tal A, Torous J. The digital mental health revolution: opportunities and risks. Psychiatr Rehabil J 2017 Sep;40(3):263-265. [doi: 10.1037/prj0000285] [Medline: 28891658]

88. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]

89. Jaafari N, Frasca M, Rigalleau F, Rachid F, Gil R, Olié J, for Insight Study Group. Forgetting what you have checked: a link between working memory impairment and checking behaviors in obsessive-compulsive disorder. Eur Psychiatry 2013 Feb;28(2):87-93. [doi: 10.1016/j.eurpsy.2011.07.001] [Medline: 21924871]

90. Lemon C, Liu N, Lane S, Sud A, Branley J, Khadra M, et al. Changes in user perceptions of a telemedicine system over time: from initial implementation to everyday use. Telemed J E Health 2018 Jul;24(7):552-559. [doi: 10.1089/tmj.2017.0194] [Medline: 29261033]

Abbreviations

AI: artificial intelligence
BT Steps: computer-driven telephone interview system
CBT: cognitive behavioral therapy
cCBT: computerized cognitive behavioral therapy
CGI: Clinical Global Impression scale
CY-BOCS: Children's Yale-Brown Obsessive-Compulsive Scale
ECG: electrocardiogram
EEG: electroencephalogram
eHealth: electronic health
EMA: ecological momentary assessment
EMG: electromyogram
ERP: exposure and response prevention
iCBT: Web-based cognitive behavioral therapy
ML: machine learning
MRI: magnetic resonance imaging
NF: neurofeedback
OCD: obsessive-compulsive disorder
OFC: orbitofrontal cortex
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
SF: sham feedback
SSRI: selective serotonin reuptake inhibitors
SVM: support vector machine

SVR: support vector regression
TA-CBT: therapist-administered cognitive behavioral therapy
tCBT: technology empowered cognitive behavioral therapy
VR: virtual reality
Y-BOCS: Yale-Brown Obsessive-Compulsive Scale

XSL•FO
RenderX

Original Paper

Identifying Sleep-Deprived Authors of Tweets: Prospective Study

Sara Melvin[1], MSc; Amanda Jamal[2], MD; Kaitlyn Hill[3], MD; Wei Wang[4], PhD; Sean D Young[5,6], PhD

[1]University of California Institute for Prediction Technology, Los Angeles, CA, United States

[2]Department of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[3]New York University-Winthrop Hospital, Mineola, NY, United States

[4]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, United States

[5]Department of Medicine, University of California, Irvine, Orange, CA, United States

[6]University of California Institute for Prediction Technology, Irvine, CA, United States

Corresponding Author:
Sean D Young, PhD
Department of Medicine
University of California, Irvine
333 City Blvd West, Suite 640
Orange, CA
United States
Phone: 1 310 456 5239
Email: syoung5@uci.edu

Abstract

Background: Social media data can be explored as a tool to detect sleep deprivation. First-year undergraduate students in their first quarter were invited to wear sleep-tracking devices (Basis; Intel), allow us to follow them on Twitter, and complete weekly surveys regarding their sleep.

Objective: This study aimed to determine whether social media data can be used to monitor sleep deprivation.

Methods: The sleep data obtained from the device were utilized to create a tiredness model that aided in labeling the tweets as sleep deprived or not at the time of posting. Labeled data were used to train and test a gated recurrent unit (GRU) neural network as to whether or not study participants were sleep deprived at the time of posting.

Results: Results from the GRU neural network suggest that it is possible to classify the sleep-deprivation status of a tweet's author with an average area under the curve of 0.68.

Conclusions: It is feasible to use social media to identify students' sleep deprivation. The results add to the body of research suggesting that social media data should be further explored as a potential source for monitoring health.

Introduction

Background

Sleeping fewer than 7 hours per night (ie, sleep deprivation) has been associated with a large number of public health concerns, including elevated blood pressure, weight gain, impaired glucose tolerance, type 2 diabetes mellitus, increased anxiety levels, and cardiovascular disease [1-6]. Poor sleep has also been associated with cognitive and motor performance deficiencies, which can lead to car accidents [7], plane crashes [8,9], and medical errors [10]. Unfortunately, the prevalence of

sleep deprivation has increased by 31% (31/100) from 1985 to 2012 [11]. In 2014, 34.8% (348/1000) of US adults received, on average, 6 hours of sleep or less in a 24-hour period [12].

Sleep deprivation is difficult to measure because of limited measurement tools. Owing to the difficulty in recruiting participants to sleep in a sleep laboratory for long periods where they can be objectively studied, most sleep studies typically use self-reported items that carry subjective bias [13,14]. For example, according to the Behavioral Risk Factor Surveillance System, during 2009-2010, an estimated 1 in 25 adult drivers (aged 18 years or older) reported having fallen asleep while

XSL·FO
RenderX

driving in the previous 30 days, suggesting limitations in self-reports of tiredness as people are not aware of their tiredness state or the impact it may have on their performance [15,16]. Although more objective technological advancements such as actigraphy and home polysomnography are available for use in research, there are limitations in applying these research-grade devices in large epidemiological and real-world settings. Therefore, new methods and tools are needed to help determine whether and when a person is sleep deprived.

Consumer-wearable smart watches and social media might be able to be used as an easy-to-integrate and more objective measure to monitor whether people are sleep deprived [17]. Unlike research-grade sleep-tracking devices, consumer wearables have the potential to monitor sleep in an unobtrusive way as consumers are naturally wearing them in daily life. However, a limitation of wearable devices is that people often choose to not wear them, thereby reducing the ability to gain sleep-tracking data.

In addition, one could use mobile phone to track their sleep; however, for many mobile phone sleep apps to work successfully, the phone has to stay on the bed the whole night to measure, there can only be one person in the bed, and the phone has to have sufficient amount of power to run an app throughout the night. Therefore, social media, which has been found to be useful as a tool for remote monitoring of behaviors, might be able to address this limitation and be used as an additional sleep monitoring tool. For example, researchers have already shown it is possible to mine text data within social media sites such as Facebook and Twitter to monitor and predict health outcomes, such as foodborne illness, influenza outbreaks, and HIV [18-20], and to monitor health behaviors [21-25]. Social media text might be similarly mined and studied to monitor sleep deprivation.

In addition, studies support that language skills appear to be affected by sleep deprivation, providing further support to our hypothesis that social media text data could be used to identify sleep deprivation [26]. For example, Harrison and Horne showed that sleep-deprived individuals generate fewer words and use less novel word associations when performing a word association task [27]. Therefore, it was hypothesized that linguistic features could passively and nondirectly characterize the tiredness state of the author at the time they created the text.

Objective

This study sought to determine the feasibility of integrating wearable smart watches and social media to monitor and verify sleep deprivation among first-year students in college. We further explored whether wearable device data and social media data could be used as tools for remotely monitoring *tiredness*. We hypothesized that students would tweet differently when they were sleep deprived compared with when they were not and that tweet data could therefore be used as a method for identifying sleep deprivation among students. Our method leverages the use of machine learning in sleep-deprivation linguistic characteristics in digital communications [27]. Machine learning algorithms have been used in several areas of health [20]; however, these methods have not been previously applied to sleep research.

Methods

Study and Participants

In 2015, between October and December, 197 first-year undergraduate students from the University of California Los Angeles (UCLA) enrolled to participate in a study aimed to analyze sleep and stress patterns among university undergraduates. Students were targeted for this study because they were a convenience sample and would provide the necessary data; 94.1% (941/1000) of college students use social networking sites [28], and the average college student spends 94.6 min per day doing various mobile phone activities, such as checking their social media networks, texting, and checking or sending emails [29].

To qualify for the study, students had to meet the following criteria: be 18 years of age or older and younger than 21 years of age, be a first-year or first-year transfer student, be in their first semester at UCLA, and have at least three posts per week on Twitter. Students self-reported these criteria, and then a research assistant verified their student status through their student ID card and verified their Twitter use by accessing their Twitter profile. The participants allowed us to follow them on Twitter to collect their tweets during the course of this study. Subjects were provided US $5 for each completed survey and an additional US $5 if all surveys were completed in a month. The total was disbursed to students after the study ended in the form of an Amazon gift card. The UCLA institutional review board approved the study protocol.

Students were asked to wear an Intel Basis sleep monitoring device, allow us to follow them on Twitter, and complete weekly Web-based surveys to self-assess psychological and sleep health, including sleep quality, stressors that week, ability to deal with these stressors, and their emotions. Owing to the subjective nature of sleep deprivation [30], several scales have been used in the clinical and scientific community to clarify the definition of sleep deprivation. These surveys assessed perceived sleep deprivation by determining the quality of sleep during the previous week and the previous night on a 5-point Likert scale of very bad, bad, average, good, and very good. Out of the 197 students who originally signed up for the study, 86 students tweeted at least once, took at least one survey, and got at least one read on their smart wristband. However, only 64 students consistently tweeted every week throughout the study. All 86 student tweets had 17,889 unique words where the average word (which includes URLs, hashtags, and mentions) length was 10.023 (SD 6.416).

Finally, participants' tweets were gathered while they were enrolled in the study using Twitter representational state transfer application program interface.

Data Classification

The goal of this study was to use Twitter data to create a model to classify whether a tweet was made by a person who was sleep deprived at the time it was posted. To accomplish this goal, we first had to develop a data processing method to properly label every tweet as to whether the author was sleep derived at the

time of the post or not. Then, a model had to be trained to classify these tweets to their correct category.

Data Preprocessing

The concept of *tiredness* is a complex notion that rises and falls throughout the day depending on a variety of factors such as quantity of sleep the night before. The Intel Basis bands provided minute-level sleep-tracking data, allowing us to use it to estimate an initial model of how tired a person is throughout their day. On the basis of the work by Pressman [31], we defined sleep deprivation as sleeping for fewer than 6 hours within 24 hours.

A sleep-labeling algorithm was created based on a simple linear model (f(x)=mx+b), where the start of every line begins at the end of a *new-day sleep*. A *new-day sleep* refers to the sleep duration that starts on one day and ends on the next day, or it starts the day after the last new-day sleep ended. All nap durations that were within that new-day plus the new-day sleep total duration were combined for the total estimate of sleep duration after the new-day sleep. Therefore, a student with a total amount of sleep greater than 6 hours (360 min) at the end of a new-day sleep was seen as starting out their day with a tiredness level (TL) of zero. This resetting is based on Pressman's [31] work on sleep deprivation and is a simplification of the real-world tiredness model; however, our method will show to be sufficient for our needs. Any other duration of sleep less than 6 hours started out with a TL of 360-γ, where γ represents the total amount of the student's sleep in minutes. Therefore, the equation to determine the minimum TL after a new-day sleep is max (360-γ,0).

A simple linear model is used to describe a person's cognitive ability as a person grows more sleep deprived. This type of model was used because Dawson and Reid [32] showed a form of linear digression of a person's cognitive performance over a period, where at 16 hours, it is equivalent to the performance of a person with a 0.02 g/dL blood alcohol level, which, for context, is greater than the US California blood alcohol limit for a person younger than 21 years. Finally, the TL threshold for sleep deprivation is considered 360 because if a person gets the minimum amount of sleep (ie, 1 min of sleep), then they will start out with the maximum TL and will be considered sleep deprived (SLD). Therefore, the slope of every linear segment is then $m=(max_{TL}-min_{TL})/(t_{SLD}-t_{wake})$ where max_{TL} is the maximum TL after a subject has been awake for over 16 hours after a perfect new-day sleep, min_{TL} is the minimum TL after a perfect new-day sleep, t_{SLD} is the amount of time it takes, in minutes, for a person to become sleep deprived after a perfect amount of sleep, and t_{wake} is the amount of time, in minutes, that has lapsed since the start of the new-day sleep. The min_{TL} after a perfect night's sleep will always be zero. In addition, the time lapsed after a person just woke up will always be zero as well. Therefore, every linear segment will have the slope of $m=(max_{TL}-min_{TL})/(t_{SLD}-twake)=360/(16*60)=3/8$.

For example, as portrayed in Figure 1, a student sleeps a full 7 hours. As this sleep is the first sleep, we assume that it is the *new-day* sleep and start the tiredness model's linear segment when the student wakes up that morning at 07:00. The student received 7 hours of sleep (420 min) the first night; thus, the y-intercept variable in f(x)=(3/8) x+b is max(360-420,0)=0. After 16 hours, the student is considered sleep deprived until they fall asleep for another new-day sleep, which approximately happens between the hours of 23:00 and 00:00 in this example.

Figure 1. This shows a numeric model example of a student's tiredness throughout their day. To label student tweets at any given minute, a simple tiredness model was created to help define when an author of a post was sleep deprived and when they were sleep sufficient. If a tweet was posted during the time a person was in their sleep deprived state (during the moment in time that the linear segment was above a tiredness level of 360), those tweets were labeled as sleep deprived while all others were labeled as sleep sufficient.
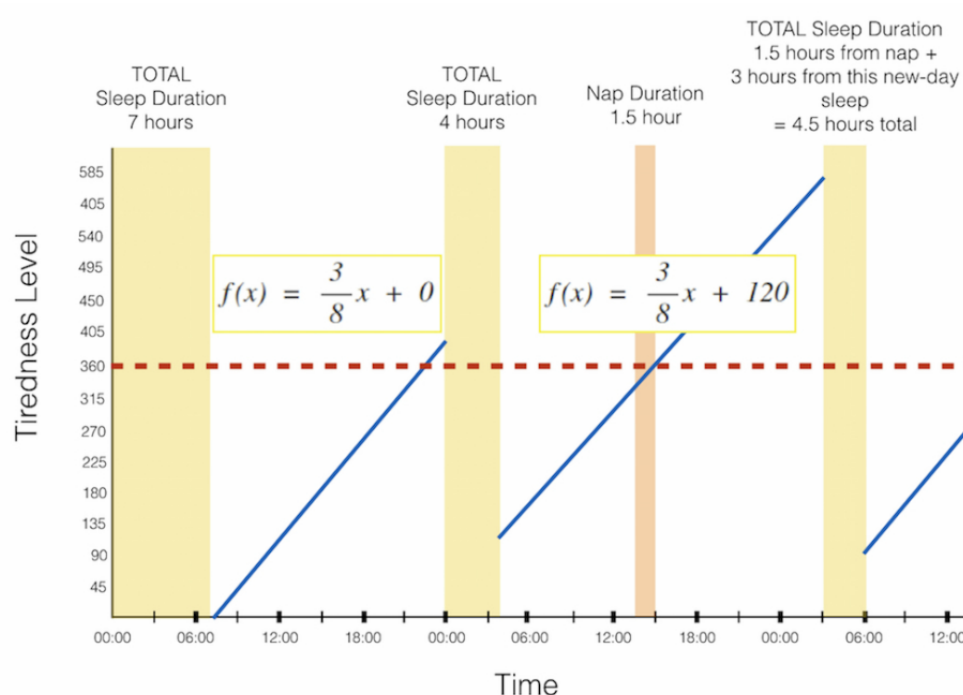
Any tweets during the student's sleep-deprived time will be labeled as *sleep deprived*, whereas all tweets that occurred between the hours of 07:00 and 23:00 will be labeled as *sleep sufficient*.

The start of the student's next sleeping period occurs on the next day after the student wakes up from their last new-day sleep. Therefore, the second sleeping period shown in this example is considered a new-day sleep. The student, in this example, received only 4 hours of sleep (240 min), thus the y-intercept variable in $f(x)=(3/8) x+b$ is max(360-240,0)=120. The student did not sleep enough that night, so they took a nap around 13:30 for approximately 1.5 hours. As this sleep does not start a day after the last new-day sleep occurred nor does this sleeping period end in the next day, this period of rest is considered a nap and will be added to the following new-day sleep.

Therefore, the time this student was sleep deprived, based upon our definition, would be between 14:40 and 03:00 the next day, when the student started their next new-day sleep. Any tweets by that student between that period will be labeled as *sleep deprived*.

Data Classification Model

After the data are labeled, we can then use supervised models to classify whether a student is sleep deprived or not based upon their Twitter posts. Out of all supervised methods, gated recurrent unit (GRU) [33], a type of recurrent neural network,

was chosen to classify tweets based upon its abilities to take a tensor as an input and to consider the ordering of words into the calculation of the final classifier [34]. These key unique characteristics of the GRU are what aided this model's ability to prove its superior performance compared with other supervised methods.

For most supervised methods, the input must be a matrix in the domain $\mathbb{R}^{n \times m}$. Originally, this study created an input matrix in the domain $\mathbb{R}^{n \times m}$ where each row represents a tweet and each element (ie, word) in the row is represented by the bag-of-words method. The bag-of-words model counts frequency of terms and does not consider the order of the words or their similarity in meaning to other words. Therefore, to add correspondence of word meaning while also maintaining word ordering, the input matrix was extended to an input tensor. Similar to the input matrix in the domain $\mathbb{R}^{n \times m}$, each row represents a tweet. However, instead of a bag-of-words term frequency representation of a word, each word is represented by a word vector, and the ordering of the words in the sentence is maintained. Hence, the input was represented by a tensor I in the domain $\mathbb{R}^{\delta \times \kappa \times \xi}$ where δ is the number of tweets, κ is the maximum tweet length of all the posts in the dataset, and ξ is the length of the word vectors used plus 3. The additional 3 is to account for the concatenation of the hour of the post, the sentiment of the post, and subjectivity of post integers to the word vector (see Figure 2).

Figure 2. Input tensor description of each dimension. Each row of the input tensor represents a tweet while each column of the tensor is a word in the tweet (not including stopwords or non-English characters). Lastly, the third dimension is the numeric vector representation of a word concatenated with the sentiment of the tweet, the subjectivity of the tweet, and the hour the tweet was posted.
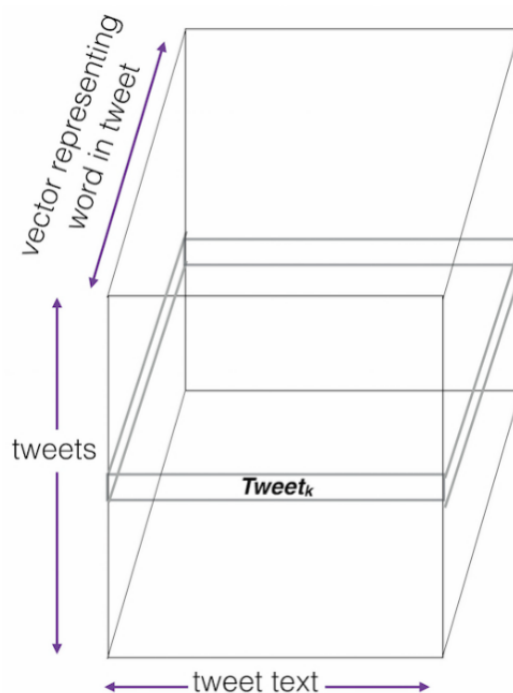
To elaborate further, every row of the input matrix I represents a tweet without stop words (ie, words that give no meaning such as *the* and *a*) and non-English characters and each word in that tweet was represented by a continuous bag-of-words (CBOW) word vector [35] that extends along the third dimension (see Figure 3). The next layer in the third dimension was the sentiment polarity where the range (0,+1) is a positive tweet, (−1,0) is a negative tweet, and 0 is a neutral tweet. In addition,

the next layer is the tweet's subjectivity that ranges from 0 to 1 where 0 is very objective and 1 is very subjective. Both the sentiment and the subjectivity were calculated using the TextBlob sentiment analysis library [36]. The last layer, the third dimension, is the hour of the tweet post in a 24-hour representation (HH). Finally, if a tweet has fewer words within it than κ, then the rest of the row is filled with zeros.

Figure 3. Example tweet converted to the matrix format that represents two dimensions of the input tensor. This example will represent what will be the slice of our input tensor labeled as tweet$_k$ in Figure 2. The stopwords (ie, words that have no meaning such as the and a) are removed thus only three words remain. Each word has a word vector associated with it from a trained continuous bag-of-words model. Lastly, word vector representation is concatenated with a holistic tweet feature: the sentiment polarity, subjectivity, and hour of post.
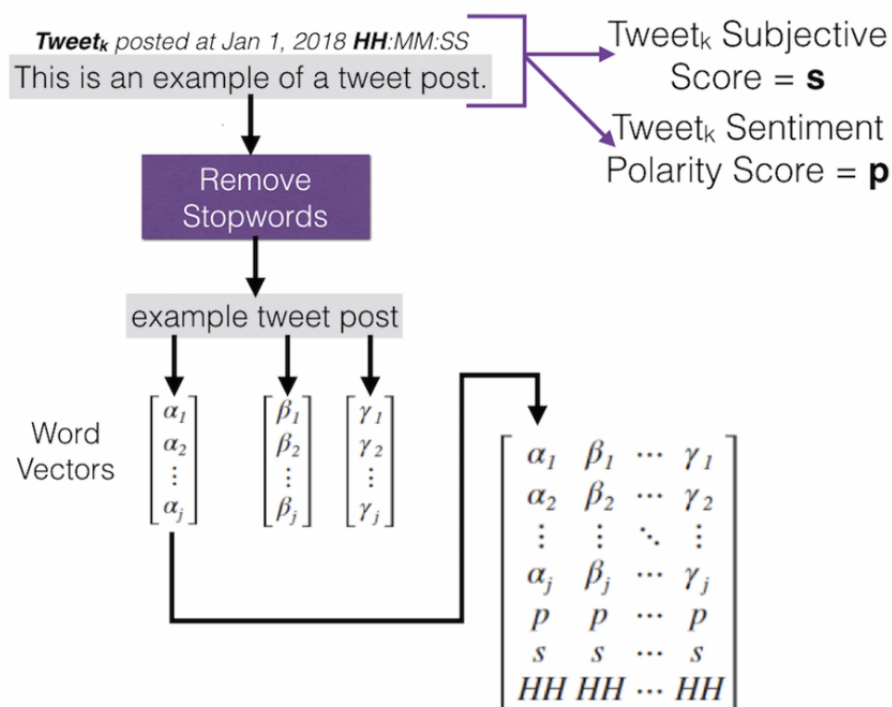
Results

Experiment Setup and Data

All tweets in this study were gathered from the UCLA first-year student study and were labeled using the simple linear model created from their Basis band data (described in the Data Preprocessing subsection). Each word in the UCLA first-year students' tweets was associated with a word vector; however, the freshmen Twitter dataset was not large enough to create a semantically accurate word vector representation. Therefore, all word vectors were trained on a larger Twitter dataset from the University of California, Irvine [37], covering a duration of 7 months and consisting of 720 million tweets, 6.6 billion words, and 3.9 million unique words. The word vectors were trained using the CBOW Word2Vec algorithm, with a dimension of 100 and a window of length 5. In addition, rare words, occurring fewer than 5 times, were removed.

The classification algorithm had 18,394 labeled UCLA freshmen's tweets, where 8,068 were sleep deprived tweets and 10,326 were sleep-sufficient tweets. The maximum sentence size with Spanish and English stop words removed was 34. The training/validation and testing dataset were split 90/10 and the

training and validation dataset were split from the previous 90 into another 90/10 split. Finally, the GRU's dense layer had a dropout rate of 0.5 for generalization purposes.

To keep from misclassifying tweets, it was assumed that a participant must sleep within 28 hours of their last new-day sleep. This threshold of 28 hours was determined based on the 2017 Accreditation Council for Graduate Medical Education regulation of a maximum shift length. This threshold seemed reasonable because studies are performed to create this threshold for medical students, and it was assumed that a freshman student would not stay awake (even if it is only for a 15-min nap) any longer than this maximum shift length. Therefore, if the sensor exhibited a student staying awake for over 28 hours, this student is showing signs of missing data, and every tweet after that 28th hour and before the next new sleep is removed from the dataset.

Experiment Results

The final metric of success was calculated using the area under the receiver operating characteristic curve (AUC) to get a fair comparison of the imbalanced dataset where 0.5 is a performance metric as good as random. With the above parameters, the GRU showed the best results, with an average AUC of 0.68 with 0.003 SD. This result shows promise that it

is possible to identify when a student is sleep deprived based solely on their Twitter posts.

Discussion

Principal Findings

The findings suggest the feasibility of integrating wearable device data and social media data to monitor sleep deprivation.

We identified 2 key points from this study. First, for researchers to be able to use wearable devices as a method of monitoring tiredness, models need to be created that can use wearable devices to monitor sleep quality and quantity to identify when a person is at a high TL to be considered sleep deprived. As, to our knowledge, there are no current numerical models tracking a person's real-time levels of tiredness, there is a lack of ground truth determining when a person is sleep deprived. This study defines tiredness using a simple linear model based on the student's Intel Basis band data to create data labels for the student's tweets. Future researchers can build on this attempt to model tiredness to more accurately identify TLs.

This study is one of the first to integrate multiple sources of remote data, including social media, self-reported Web-based surveys, and Intel Basis band data. We integrate these different data sources to develop a tiredness model and to passively monitor sleep deprivation from nonsensor devices such as social media. Results suggest the feasibility of using wearable smart watches and social media (ie, Twitter) data for monitoring sleep deprivation among undergraduate students. We found, in our study, that we were able to train and test a model that used Twitter data and could predict student's objective level of sleep, as measured by a sleep monitoring device. Therefore, because of the popularity of social media and the predicted accuracy of the results, a novel technique to assess global sleep sufficiency and deprivation has been shown.

Limitations and Future Work

Although our early results look promising, there are limitations to this study. First, because of the requirements that the participating student must have worn their sleep-tracking device, taken a Web-based survey, and tweeted at least 3 times per week, this pilot feasibility study was limited to a final sample size of 86 out of 197 freshmen students. Second, data were only recorded during the fall semester, thus limiting this classification model to a specific demographic in a specific time window. In the future, research can expand upon this work to include an entire school year and across several other universities to obtain more generalizable sleep-deprivation results.

In addition, we recognize that student populations have different behaviors and related sleep factors (eg, feeling the need to stay up all night to study for tests) that can affect tiredness and sleep deprivation differently from other occupations, thus reducing the ability to generalize these findings to all people and populations.

A further limitation to this study is that we used Twitter as the only social media platform. Participants were screened to include those who frequently use Twitter because Twitter is a frequently used data source in modeling research. It is possible that people tweet differently than they use Instagram or Facebook, so there may be signs of sleep deprivation that were missed because other social media platforms were not used in this study. This question can be studied in future research and has been added as a recommendation for future research.

It is also noted that it is possible that students changed their Twitter behavior because they had knowledge of being observed. Therefore, another limitation in this study is that we have no way to know whether a student varied their language based on being observed or not; as a result, it is assumed that the language was not altered. The study of how sleep-deprived and sleep-sufficient language changes when being observed or not observed is left for future research.

In addition, more research needs to be performed in verifying the device and identifying its average error as we used bands provided to us and did not conduct validation methods to assess the accuracy of the readings for each band. Furthermore, there was no metric available to determine how tired a person is throughout every minute of their day; therefore, a simplistic model was created to best define a student's tiredness based upon their previous night's sleep. This model limits the accuracy of the classification model through the numeric model's rough approximation of the sleep-deprived period. Similarly, sleep deprivation is subjective (ie, people have differing levels of sleep needed), making it difficult to evaluate sleep deprivation objectively. We attempted to address this issue by validating the study participant data by corresponding one night or less a week wearable (objective) data to a weekly survey (subjective) measurement asking how the student slept the night before. Future research can build and improve on these methods for more accurate classification.

Finally, the numeric model was limited to only including sleep quantity to determine the beginning tiredness/sleep-deprivation level of the student in the morning. This model could be further improved by incorporating sleep quality, types of beverages, and food consumed throughout the day; types of medications taken; and exercise duration and type into the model. For determining tiredness at the beginning of the student's day, sleep quality could be quantitatively measured using data captured by smart wearables. Studies have shown that the number of interruptions in sleep and the percentage of each sleep stage throughout the night determine the sleep quality [38]. However, more research needs to be performed to numerically model sleep quality and determine how it affects the TL variance throughout the day.

Conclusions

This pilot study suggests the feasibility of (1) modeling a student's TL throughout the duration of their day from smart wristband devices and (2) determining whether a student is sleep deprived based on their social media behavior. Future research should further explore the integration of multiple data sources to monitor real-time changes in tiredness.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Lu H, Croft JB. Prevalence of healthy sleep duration among adults--United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Feb 19;65(6):137-141 [FREE Full text] [doi: 10.15585/mmwr.mm6506a1] [Medline: 26890214]

2. van Cauter E, Knutson KL. Sleep and the epidemic of obesity in children and adults. Eur J Endocrinol 2008 Dec;159(Suppl 1):S59-S66 [FREE Full text] [doi: 10.1530/EJE-08-0298] [Medline: 18719052]

3. Mullington JM, Haack M, Toth M, Serrador JM, Meier-Ewert HK. Cardiovascular, inflammatory, and metabolic consequences of sleep deprivation. Prog Cardiovasc Dis 2009;51(4):294-302 [FREE Full text] [doi: 10.1016/j.pcad.2008.10.003] [Medline: 19110131]

4. Nagai M, Hoshide S, Kario K. Sleep duration as a risk factor for cardiovascular disease- a review of the recent literature. Curr Cardiol Rev 2010 Feb;6(1):54-61 [FREE Full text] [doi: 10.2174/157340310790231635] [Medline: 21286279]

5. Pires GN, Bezerra AG, Tufik S, Andersen ML. Effects of acute sleep deprivation on state anxiety levels: a systematic review and meta-analysis. Sleep Med 2016 Aug;24:109-118. [doi: 10.1016/j.sleep.2016.07.019] [Medline: 27810176]

6. Spiegel K, Knutson K, Leproult R, Tasali E, van Cauter E. Sleep loss: a novel risk factor for insulin resistance and Type 2 diabetes. J Appl Physiol (1985) 2005 Nov;99(5):2008-2019. [doi: 10.1152/japplphysiol.00660.2005] [Medline: 16227462]

7. US Department of Transportation, National Highway Traffic Safety Administration. CrashStats - NHTSA - DOT. 2008 Jul. National Motor Vehicle Crash Causation Survey: Report to Congress URL: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811059 [accessed 2019-11-04]

8. Federal Aviation Administration. 2014 Jul 18. NASA Controller Fatigue Assessment Report URL: https://www.faa.gov/data_research/research/media/nasa_controller_fatigue_assessment_report.pdf [accessed 2019-11-04]

9. Thompson DE. Surveying Air Traffic Control Specialist Perception of Scheduling Regulations. In: A Thesis Presented to the Faculty of the College of Basic and Applied Sciences Middle Tennessee State University in Partial Fulfillment of the Requirements for the Degree Master of Science in Aviation Administration. Murfreesboro, TN: MS thesis, Middle Tennessee State University; 2015.

10. Rogers AE, Hwang W, Scott LD, Aiken LH, Dinges DF. The working hours of hospital staff nurses and patient safety. Health Aff (Millwood) 2004;23(4):202-212. [doi: 10.1377/hlthaff.23.4.202] [Medline: 15318582]

11. Ford ES, Cunningham TJ, Croft JB. Trends in self-reported sleep duration among US adults from 1985 to 2012. Sleep 2015 May 1;38(5):829-832 [FREE Full text] [doi: 10.5665/sleep.4684] [Medline: 25669182]

12. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Hua L, Croft JB. Prevalence of Healthy Sleep Duration among Adults — United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Feb 19;65(6):137-141 [FREE Full text]

13. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]

14. Dewald JF, Meijer AM, Oort FJ, Kerkhof GA, Bögels SM. The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. Sleep Med Rev 2010 Jun;14(3):179-189. [doi: 10.1016/j.smrv.2009.10.004] [Medline: 20093054]

15. Wheaton A, Chapman D, Presley-Cantrell L, Croft J, Roehler D. Drowsy driving — 19 states and the District of Columbia, 2009–2010. MMWR Morb Mortal Wkly Rep 2013 Jan 4;61(51 & 52):1033-1037 [FREE Full text]

16. Wheaton A, Shults R, Chapman D, Ford E, Croft J. Drowsy driving and risk behaviors — 10 states and Puerto Rico, 2011–2012. MMWR Morb Mortal Wkly Rep 2014 Jul 4;63(26):557-562 [FREE Full text]

17. Ferrie JE, Kumari M, Salo P, Singh-Manoux A, Kivimäki M. Sleep epidemiology--a rapidly growing field. Int J Epidemiol 2011 Dec;40(6):1431-1437 [FREE Full text] [doi: 10.1093/ije/dyr203] [Medline: 22158659]

18. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 2012 Jan;86(1):39-45 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0597] [Medline: 22232449]

19. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PLoS One 2011 May 4;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]

20. Young SD, Yu W, Wang W. Toward automating HIV identification: machine learning for rapid identification of HIV-related social media data. J Acquir Immune Defic Syndr 2017 Feb 1;74(Suppl 2):S128-S131 [FREE Full text] [doi: 10.1097/QAI.0000000000001240] [Medline: 28079723]

21. McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions - evidence from Twitter. J Am Med Inform Assoc 2017 May 1;24(3):496-502. [doi: 10.1093/jamia/ocw133] [Medline: 27707822]

22. Fisichella M, Stewart A, Cuzzocrea A, Denecke K. Detecting Health Events on the Social Web to Enable Epidemic Intelligence. In: Proceedings of the 18th international conference on String processing and information retrieval. 2011 Presented at: SPIRE'11; October 17-21, 2011; Pisa, Italy p. 87-103 URL: https://doi.org/10.1007/978-3-642-24583-1_10

23. Lejeune G, Brixtel R, Doucet A, Lucas N. Multilingual event extraction for epidemic detection. Artif Intell Med 2015 Oct;65(2):131-143. [doi: 10.1016/j.artmed.2015.06.005] [Medline: 26228941]

24. Culotta A. Towards Detecting Influenza Epidemics by Analyzing Twitter Messages. In: Proceedings of the First Workshop on Social Media Analytics. 2010 Presented at: SOMA'10; July 25-28, 2010; Washington DC p. 115-122 URL: https://doi.org/10.1145/1964858.1964874 [doi: 10.1145/1964858.1964874]

25. Aramaki E, Maskawa S, Morita M. Twitter Catches the Flu: Detecting Influenza Epidemics Using Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: EMNLP'11; July 27-31, 2011; Edinburgh, Scotland, UK p. 1568-1576 URL: https://www.aclweb.org/anthology/D11-1145

26. Odle-Dusseau HN, Bradley JL, Pilcher JJ. Subjective perceptions of the effects of sustained performance under sleep-deprivation conditions. Chronobiol Int 2010 Jan;27(2):318-333. [doi: 10.3109/07420520903502226] [Medline: 20370472]

27. Harrison Y, Horne JA. Sleep loss impairs short and novel language tasks having a prefrontal focus. J Sleep Res 1998 Jun;7(2):95-100 [FREE Full text] [doi: 10.1046/j.1365-2869.1998.00104.x] [Medline: 9682180]

28. Smith S, Caruso J. Anita Crawley. 2010 Oct. ECAR Study of Undergraduate Students and Information Technology, 2010 URL: http://anitacrawley.net/Resources/Reports/ECAR%20study%20highlights.pdf [accessed 2019-11-04]

29. Roberts JA, Yaya LH, Manolis C. The invisible addiction: cell-phone activities and addiction among male and female college students. J Behav Addict 2014 Dec;3(4):254-265 [FREE Full text] [doi: 10.1556/JBA.3.2014.015] [Medline: 25595966]

30. Harvey AG, Stinson K, Whitaker KL, Moskovitz D, Virk H. The subjective meaning of sleep quality: a comparison of individuals with and without insomnia. Sleep 2008 Mar;31(3):383-393 [FREE Full text] [doi: 10.1093/sleep/31.3.383] [Medline: 18363315]

31. Pressman MR. somepomed. 2011. Definition and consequences of sleep deprivation URL: https://somepomed.org/articulos/contents/mobipreview.htm?11/34/11823/abstract/14 [accessed 2017-05-05] [WebCite Cache ID 74V5U3q8T]

32. Dawson D, Reid K. Fatigue, alcohol and performance impairment. Nature 1997 Jul 17;388(6639):235. [doi: 10.1038/40775] [Medline: 9230429]

33. Cho K, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 26–28, 2014; Doha, Qatar p. 1724-1734 URL: https://doi.org/10.3115/v1/D14-1179 [doi: 10.3115/v1/d14-1179]

34. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: The Mit Press; 2016.

35. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv 2013 Jan 16 [FREE Full text]

36. Loria S. GitHub. TextBlob URL: https://github.com/sloria/TextBlob [accessed 2017-05-05] [WebCite Cache ID 74V5XeU9z]

37. Jia J, Li C, Zhang X, Li C, Carey MJ, Su S. Towards Interactive Analytics and Visualization on One Billion Tweets. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2016 Presented at: SIGSPACIAL'16; October 31 - November 3, 2016; Burlingame, California URL: http://doi.acm.org/10.1145/2996913.2996923 [doi: 10.1145/2996913.2996923]

38. Martin SE, Wraith PK, Deary IJ, Douglas NJ. The effect of nonvisible sleep fragmentation on daytime function. Am J Respir Crit Care Med 1997 May;155(5):1596-1601. [doi: 10.1164/ajrccm.155.5.9154863] [Medline: 9154863]

Abbreviations

AUC: area under the receiver operating characteristic curve
CBOW: continuous bag of words
GRU: gated recurrent unit
SLD: sleep deprived
TL: tiredness level
UCLA: University of California Los Angeles

XSL•FO
RenderX

<u>Original Paper</u>

A Virtual Reality Video to Improve Information Provision and Reduce Anxiety Before Cesarean Delivery: Randomized Controlled Trial

Lore Noben[1,2], MD; Simone Maria Theresia Anna Goossens[1,2], MD, PhD; Sophie Eva Marieke Truijens[1,2], PhD; Marijn Marthe Georgine van Berckel[3], MD; Christel Wilhelmina Perquin[4], MD, PhD; Gerrit Dirk Slooter[3], MD, PhD; Stefanus Johannes van Rooijen[3,5], MD, PhD

[1]Department of Obstetrics and Gynecology, Máxima Medical Center, Veldhoven, Netherlands

[2]Eindhoven MedTech Innovation Center, Eindhoven, Netherlands

[3]Department of Surgery, Máxima Medical Center, Veldhoven, Netherlands

[4]Department of Anesthesiology and Pain and Palliative Care, Máxima Medical Center, Veldhoven, Netherlands

[5]Care and Public Health Research Institute, Maastricht University Medical Center, Maastricht, Netherlands

Corresponding Author:
Lore Noben, MD
Department of Obstetrics and Gynecology
Máxima Medical Center
PO Box 7777
Veldhoven, 5500 MB
Netherlands
Phone: 31 643123997
Fax: 31 40888 8387
Email: lore.noben94@gmail.com

Abstract

Background: Anxiety levels before cesarean delivery (CD) can lead to a negative birth experience, which may influence several aspects of the woman's life in the long term. Improving preoperative information may lower preoperative anxiety and lead to a more positive birth experience.

Objective: This study aimed to determine whether a virtual reality (VR) video in addition to standard preoperative information decreases anxiety levels before a planned CD.

Methods: Women scheduled to undergo term elective CD were recruited from the outpatient clinic. They were randomized and stratified based on history of emergency CD (yes or no). All participants received standard preoperative information (folder leaflets and counseling by the obstetrician); the VR group additionally watched the VR video showing all aspects of CD such as the ward admission, operating theater, spinal analgesia, and moment of birth. The primary outcome measure was a change in score on the Visual Analogue Scale for Anxiety (ΔVAS-A) measured at admission for CD, compared with the baseline VAS-A score.

Results: A total of 97 women were included for analysis. The baseline characteristics were similar in both groups, except for a significantly higher level of education in the control group. There was no significant decrease in the VAS-A score of the women in the VR group (n=49) compared with those in the control group (n=48; ΔVAS-A=1.0; *P*=.08; 95% CI −0.1 to 2.0). Subgroup analysis for the group of women with a history of emergency CD showed a trend toward decreased preoperative anxiety, despite the small sample size of this subgroup (n=17; *P*=.06). Of the 26 participants who provided completed questionnaires, 22 (85%) in the VR group reported feeling more prepared after seeing the VR video; of the 24 participants' partners who completed the questionnaires, 19 (79%) agreed with the participants. No discomfort or motion sickness was reported.

Conclusions: A VR video may help patients and their partners feel better prepared when planning a CD. This study showed that VR does not lead to a decrease in preoperative anxiety. However, subgroups such as women with a history of emergency CD may benefit from VR videos.

Trial Registration: International Standard Randomised Controlled Trial Number (ISRCTN) 74794447; http://www.isrctn.com/ISRCTN74794447 (retrospectively registered)

Introduction

Background

A cesarean delivery (CD) is one of the most commonly performed surgeries in obstetrics, and the number of CDs performed is still increasing worldwide [1]. In 2017, 14% of all term deliveries in the Netherlands were performed via a CD, of which half were planned [2]. Data about the physical risks of the surgical procedure, such as infection and bleeding, are well known and should be part of the process of gaining informed consent. In the last few years, there is growing awareness about the psychological impact of CD [3].

The CD procedure is mostly performed under regional anesthesia, without sedatives or anxiolytics, to facilitate a conscious birth experience for the mother, prevent depression of the neonate, and promote immediate (skin-to-skin) contact between the mother and her baby [4].

Previous research demonstrated that the level of anxiety and fear of childbirth are known to be associated with the incidence of postpartum depression. Women who deliver by a CD are at risk for both increased fear of childbirth and postpartum depression [5]. It is essential to minimize preoperative anxiety for these patients because lower preoperative anxiety has been shown to lead to greater maternal satisfaction with CD and thus a more positive birth experience.

This is important because a negative birth experience is associated with serious negative long-term effects on several aspects of a woman's life, such as the relationship with her partner and the baby and delay or even avoidance of future pregnancies [6,7]. Information provision has been shown to be a key element for quality of care, as perceived by women who gave birth [8].

Partners of women report anxiety and fear related to childbirth as well, especially with respect to a CD [9,10]. Providing both patients and their partners with good-quality information for CD might reduce preoperative fear and anxiety and thus improve satisfaction and recovery. Furthermore, it might minimize the possible negative long-term sequelae of a negative birth experience for both [11].

Currently, preoperative information is provided by the treating physician and through information folders. Ideally, a life guided tour and step-by-step explanation of the course of the procedure during the day should be part of the program for preparation. However, this elaborate preparation would require valuable time from the hospital personnel, and time is nowadays scarce and expensive. Owing to sterility issues, it is impossible to have a *life guided tour* through the operating room. Video education for surgery and medical interventions, as an alternative approach for a life guided tour, has been proven to improve immediate and short-term knowledge [12]. However, a difference in general anxiety (and anxiety and satisfaction with the consent process) is not uniformly shown [12]. Therefore, new technologies such as virtual reality (VR) may be of additional value, as patients can virtually experience the operating room and be better prepared for their surgery.

Objectives

To our knowledge, this is the first study to investigate the effect of VR in addition to conventional information provision on the preoperative anxiety levels of women undergoing a planned CD. Our primary hypothesis was that adding the VR video to standard preoperative information would show a statistically significant decrease in preoperative anxiety compared with providing the standard preoperative information without a VR video. In addition, we expected a positive effect of VR on patients' levels of anxiety and patient satisfaction scores of both women and their partners. Furthermore, this study aimed to determine whether VR would be feasible to implement, without causing any harmful side effects such as motion sickness.

Methods

Video Development

A total of eight women at the outpatient clinic who were scheduled to undergo a CD were interviewed. During this interview, we asked the women to describe their feelings in general and in terms of anxiety toward the planned CD as well as the way they received information about CD, their satisfaction with completeness of the information, and any possible improvements. We also asked for their opinion of the use of video and VR as possible information tools. All these women received information from their gynecologist and the internet, and one of them searched and watched a Web-based conventional video. Four women indicated that they would see additional value of information in a 2D video, whereas five women felt the same about VR. In addition, one woman indicated that she did not want to see a 2D video but was positive about a VR video because of the possibility of *watching away*. Of the women we interviewed, four did not expect a VR video to reduce their preoperative anxiety, three did not know what the effect of a VR video would be on their preoperative tension, and one thought a VR video might reduce preoperative stress. On the basis of this information, we developed a VR video, which can be viewed in the Infor-Med app (Infor-Med BV) on a smartphone or tablet.

The 360° VR video shows all the aspects of a CD, including the admission on the ward, the operating room, placement of spinal analgesia, and the birth of the baby when the gynecologist lifts the baby above the sterile environment (Figure 1). Instructions on viewing and downloading the VR video can be found on the website [13]. The video did not show any surgical content such as the area of incision. The video was recorded from the partner's perspective in the operating room. The video ends at the ward, where the family is reunited. The video lasts 285 seconds and is narrated with a Dutch voice-over. Patients

and their partners were continuously involved during the development of the video to allow us to include their feedback on the images, text, and changing the sequence.

Figure 1. Screenshot from the 360 virtual reality video at the moment of birth.
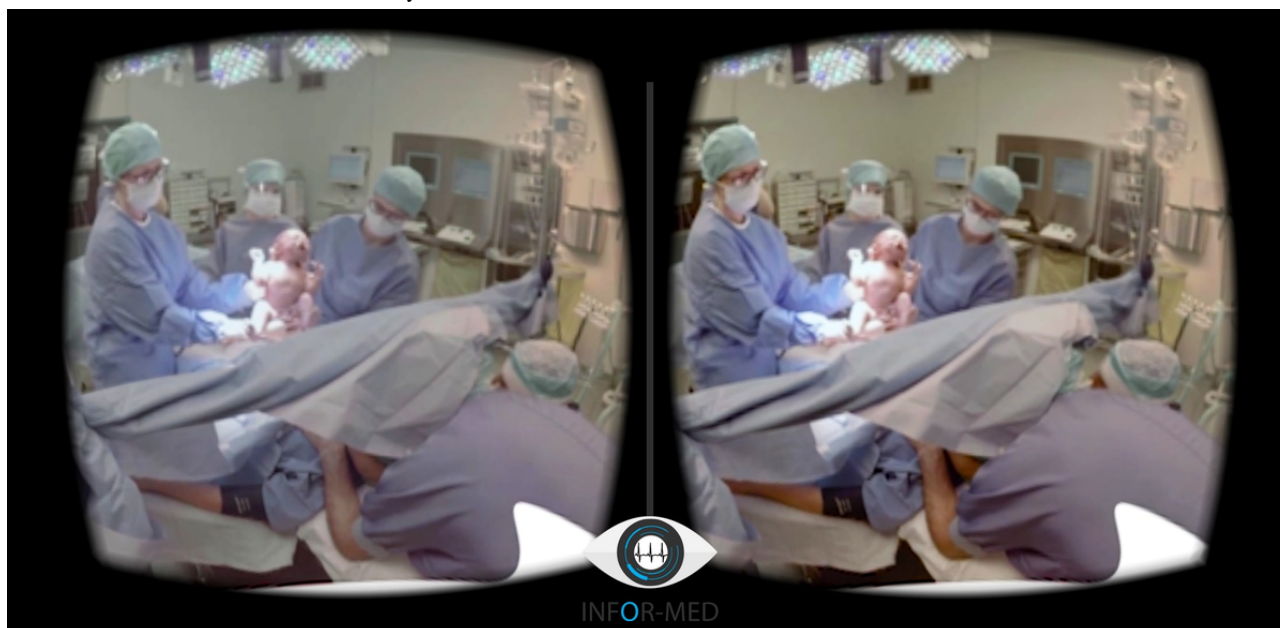
Study Design and Population

This randomized controlled trial included 80 women, enrolled from November 2016 to January 2018, who were scheduled for elective CD at Máxima Medical Center in Veldhoven, the Netherlands. We received a statement from our local institutional review board that no ethical approval was required (N17.017). Women were eligible for inclusion if they were aged 18 years or older, had planned for elective CD after 37 weeks of gestation, and had sufficient knowledge of the Dutch language. Exclusion criteria were prematurity (gestational age < 37 weeks), placenta previa, pre-eclampsia, and a suspected congenital anomaly. Patients were recruited from the outpatient clinic at our hospital. They were not explicitly informed that the study involved a VR video but were told that the intervention group received a novel method of information provision in addition to the standard information.

Randomization and Masking

Randomization was performed by the researcher (LN) using a Web-based computer randomizer generating a randomization list. Couples were randomized into two groups by means of stratified block randomization: the control group received standard information from their doctor through information leaflets and oral counseling, and the intervention group (VR group) received the standard information and an additional VR video. Randomization blocks of 10 were used. Stratification was used based on the following two subgroups: (1) women with no history of CD and (2) women with previous emergency CD. The latter group was chosen, as these women may not have received elaborate information about the procedure before their first delivery because a vaginal delivery was intended. Masking of the researcher and participants was not possible because of the nature of the intervention.

Procedures

After obtaining written informed consent, both women and their partners were asked to fill out the first questionnaire (time point 1). Subsequently, they were randomized into groups. If couples were randomized to the VR group, a VR video was shown using the Infor-Med app on the participant's smartphone, and VR glasses (Figure 2) were supplied by the researcher at the outpatient clinic. Couples received a unique password to install and watch the video at the time of inclusion or later at home (unlimited views). By using a password, we prevented the possibility of patients in the control group gaining access to the VR video. The second questionnaire was filled out on the day of CD at admission to the ward (time point 2). If the woman was admitted earlier because of contractions, the questionnaire was filled out at that time. Couples were asked to fill out the third questionnaire 1-2 weeks after CD (time point 3). The questionnaire time schedule is shown in Figure 3. The primary outcome measure was the mean score on the Visual Analogue Scale for Anxiety (VAS-A) at hospital admission (time point 2). Data on the baseline characteristics were collected from the electronic patient file.

Figure 2. Illustration of the virtual reality glasses used.

Figure 3. Time schedule and measured variables. *The intervention group additionally watched the 360° virtual reality video. £The intervention group additionally filled out the Simulation Sickness Questionnaire after watching the virtual reality video. CD: cesarean delivery; CPS: Childbirth Perception Scale; PCQ: Pregnancy and Childbirth Questionnaire; SSQ: Simulation Sickness Questionnaire; TPDS: Tilburg Pregnancy Distress Scale; VAS-A: Visual Analogue Scale for Anxiety.
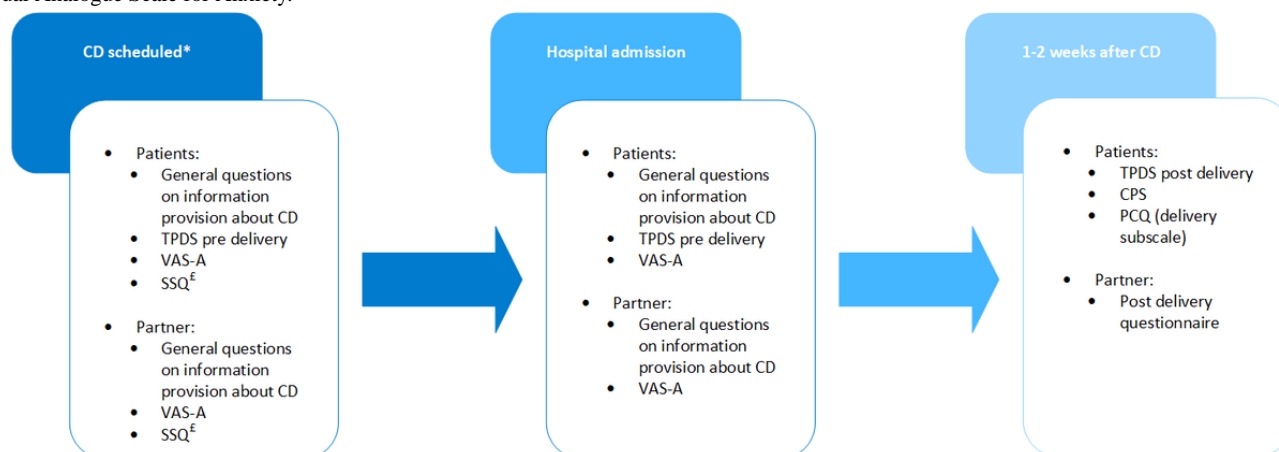
Questionnaires

The Simulation Sickness Questionnaire (SSQ) with 13 questions regarding the symptoms related to motion sickness was added to the first questionnaire for the VR group [14]. These symptoms were scaled from none (1) to severe (4).

The VAS-A was used in the first and second questionnaires to measure preoperative anxiety [15]. It comprises a 10-cm horizontal line, stating, on the left end, "not anxious at all" and, on the right end, "most anxious I can imagine."

The Tilburg Pregnancy Distress Scale was used in the first and second questionnaires. This questionnaire consists of 16 items regarding the woman's perception of her pregnancy, divided over two subscales: negative affect and partner involvement. Questions are formed in positive and negative statements. Items were recoded such that a higher score represents a higher level of distress [16]. The Childbirth Perception Scale (CPS) was used in the third questionnaire, which was filled out approximately 1 week after the CD. The CPS consists of 12 items with a *perception of delivery* subscale (6 items) and a *perception of the first postpartum week* subscale (6 items). Items were recoded such that a higher score corresponded with more

distress and less positive perception [17]. Women were also asked to fill out the *pregnancy* subscale of the Pregnancy and Childbirth Questionnaire (PCQ) in the third questionnaire, consisting of seven items. Items were recoded such that higher scores indicate higher quality of care. This third questionnaire also contained a short questionnaire with five items for the partners, regarding their experience of this postpartum week. A translated version of this short questionnaire is presented in Multimedia Appendix 1. Furthermore, we asked participants if the preoperative information was sufficient. Participants in the VR group received the additional question if they felt more prepared for CD after seeing the VR video (response: yes or no).

Statistical Analyses

Sample size calculation was performed using the software G*power3 [18] and based on VAS-A scores from the literature [19]. Given a mean VAS-A score of 5.01 (SD 3.14) cm [19], 38 patients were included in each study arm to detect a decrease in VAS-A score of 2 cm. A sample size of 38 patients per study arm was determined based on detecting a decrease in VAS-A of 2 cm, with 80% power and a significance level of .05. To account for missing data, we set the sample size at 40 inclusions per study arm.

Statistical analyses were performed using SPSS (version 25; IBM Corporation, Armonk, New York). To test for differences in baseline characteristics between the two groups, a Student *t* test and Mann-Whitney *U* test were used for normally and nonnormally distributed data, respectively. For both women and their partners, the difference in VAS-A score at the second time point between the VR group and the control group was calculated using a Student *t* test, as these data were distributed normally. To test the influence of the VR video on the change in VAS-A score between the first and second time points for each study group, a linear regression analysis was performed. Thereafter, the history of CD was added as an independent factor

in a multivariate regression analysis. The following variables were found to have an influence on the VAS-A score in previous literature: baseline VAS-A (time point 1), a psychiatric history of depression or anxiety disorder, level of education, previous CD, age, and marital status. Subsequently, randomization to the VR group was added as an independent factor in multivariate regression analysis for each variable that showed a significant contribution to the regression model.

As the history of emergency CD is known to be a predictor of preoperative anxiety, we performed a separate regression analysis to determine if there was a significant effect of the VR video in this subgroup alone. This was possible because of the stratified randomization applied.

For the same reason, differences in scores on the questionnaires were calculated separately for both the groups with and without a history of emergency CD. After recoding, the scores of the individual items were added to obtain an overall score. We calculated the difference between this overall score for each of the questionnaire subscales between the VR group and the control group using a Student *t* test.

Results

Participant Characteristics

A total of 99 patients were included. Owing to the high number of missing questionnaires at the start of the study, we decided to continue including patients until we reached 80 completed questionnaires at time point 2, which included our primary outcome measure. Moreover, two patients were excluded because of spontaneous vaginal delivery after randomization. From the remaining 97 patients, we received 94% (91/97) completed questionnaires at time point 1, 87% (84/97) completed questionnaires at time point 2, and 73% (71/97) completed questionnaires at time point 3. Figure 4 shows the flowchart of patient inclusion.

Figure 4. Flowchart of patient inclusion. CD: cesarean delivery; VR: virtual reality.
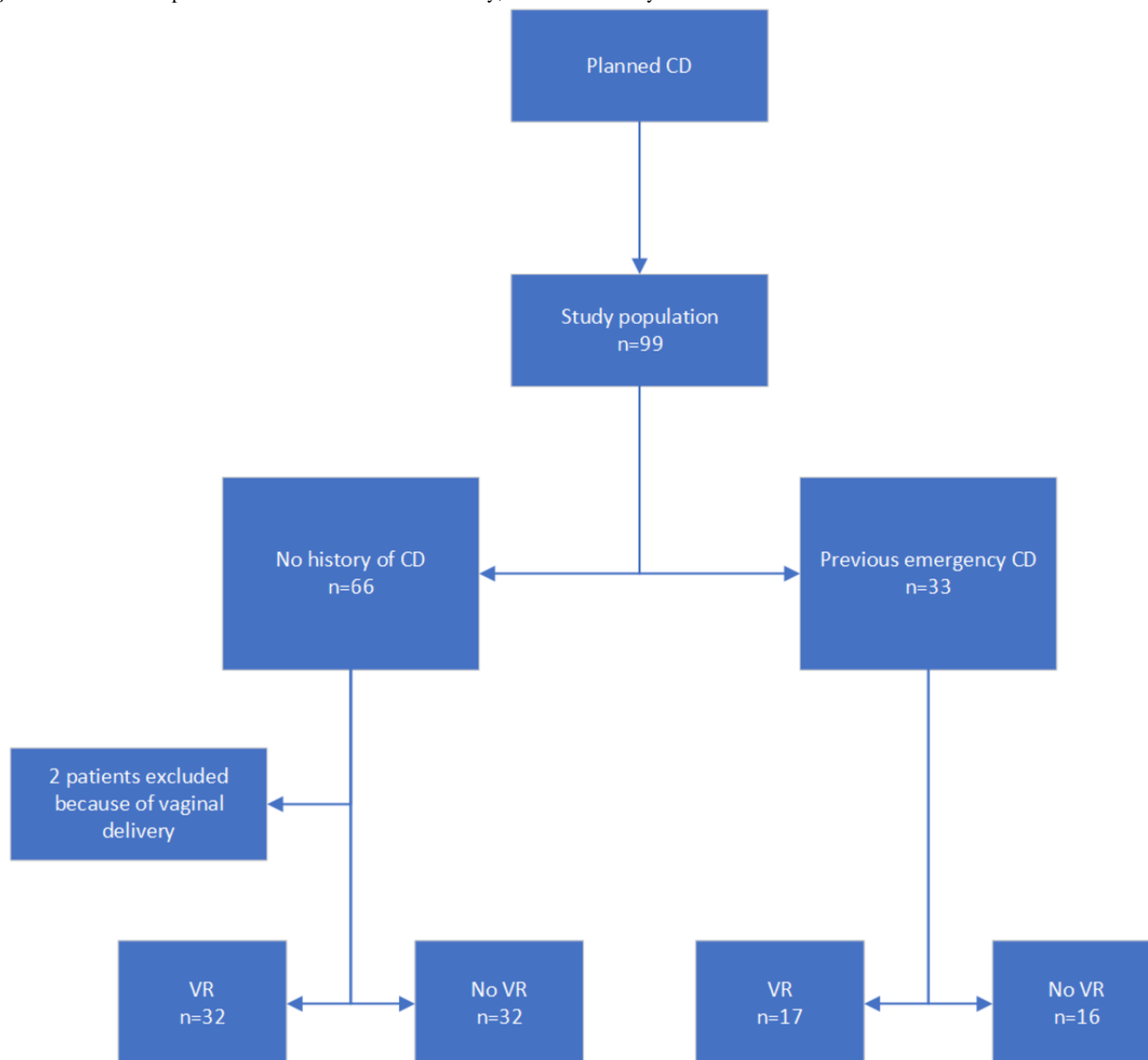
Baseline characteristics for both the VR group and the control group are shown in Table 1. There were no differences between the groups with respect to age, gestational age at delivery, parity, and the incidence of previous CD. We noticed a significant difference in the level of education, with a higher proportion of participants with a high level of education level in the control group (*P*=.03).

Table 1. Baseline characteristics of women undergoing cesarean delivery for both study groups (probability values were calculated using an independent *t* test for normally distributed data; otherwise, a Mann-Whitney *U* test was used).

Characteristics	Virtual reality group (n=49)	Control group (n=48)
Age (years), mean (SD)	32.6 (3.9)	33.12 (4.3)
Gestational age (weeks) on delivery, mean (SD)	39.0 (0.7)	38.8 (0.8)
Gravidity, median (IQR[a])	2 (1-3)	2 (1-2)
Parity, median (IQR)	1 (0-1)	1 (0-1)
Previous CD, n (%)	17 (35)	16 (33)
Marital status, n (%)		
Married	30 (61)	24 (50)
Living together	15 (31)	19 (40)
Single	0 (0)	2 (4)
Missing	4 (8)	3 (6)
Body mass index (kg/m^2), median (IQR)	24.8 (22.9-29.2)	24.9 (22.2-27.5)
History of depression or anxiety, n (%)		
Yes	13 (27)	10 (21)
No	33 (67)	35 (73)
Missing	3 (6)	3 (6)
Level of education[b], n (%)		
University/college	24 (49)	32 (67)
Secondary education	0 (0)	3 (6)
Vocational training	17 (35)	8 (17)
Prevocational education	5 (10)	2 (4)
No education or primary education	0 (0)	0 (0)
Missing	3 (6)	3 (6)
Indication for CD[c], n (%)		
Repeat CD	14 (29)	16 (33)
Fetal breech position	24 (49)	17 (35)
History of obstetric complications	5 (10)	8 (17)
Current obstetric complications	0 (0)	2 (4)
Medical history	4 (8)	2 (4)
Patient's request	2 (4)	3 (6)
Information sources, n (%)		
Health care professional	38 (78)	36 (75)
Patient folders	37 (76)	40 (83)
Experiences from friends/family	24 (49)	27 (56)
Internet	25 (51)	27 (56)
YouTube	8 (16)	10 (21)
Other	8 (16)	6 (13)

[a]IQR: interquartile range.

[b]Significant difference (*P*<.05) in ranks between the VR group and the control group as calculated with the Mann-Whitney *U* test.

[c]CD: cesarean delivery.

XSL•FO
RenderX

Primary Outcome Measures

There was no significant difference in the mean VAS-A score at admission between the control group and the VR group for both women (4.6 [SD 2.5] vs 5.6 [SD 2.4], respectively; P=.08) and their partners (3.4 [SD 2.0] vs 3.9 [SD 2.5], respectively; P=.30). There was no difference between the control group and the VR group in terms of baseline VAS-A scores for the women (3.8 [SD 2.3] vs 4.1 [SD 2.3], respectively; P=.52) or their

partners (2.5 [SD 1.9] vs 2.5 [SD 2.3], respectively; P=.98). There was an increase in the VAS-A score (ΔVAS-A) between the first and second measurements of 1.5 cm for the women in the VR group compared to 0.8 cm for women in the control group (95% CI −0.1 to 2.0; P=.08). For their partners, there was an increase of 1.4 cm in the VR group compared to 0.9 cm in the control group (95% CI −0.5 to 1.6; P=.30). Table 2 gives an overview of these results.

Table 2. Difference in the Visual Analogue Scale for Anxiety value at time point 2 (at admission) between the virtual reality and the control groups for both women and their partners (difference calculated using the Student t test).

Study arm	VAS-A[a] Q1[b], mean (SD)	VAS-A Q2[c], mean (SD)	Change in score on the Visual Analogue Scale for Anxiety	Mean difference[d]	P value[d]	95% CI[d]
Women				1.0	.08	−0.1 to 2.0
Control group	3.8 (2.3)	4.6 (2.5)	0.8			
VR[e] group	4.1 (2.3)	5.6 (2.4)	1.5			
Partners				0.6	.30	−0.5 to 1.6
Control group	2.5 (1.9)	3.4 (2.0)	0.9			
VR group	2.5 (2.3)	3.9 (2.5)	1.4			

[a]VAS-A: Visual Analogue Scale for Anxiety.

[b]Q1: questionnaire time point 1.

[c]Q2: questionnaire time point 2.

[d]These values are given for the change in score on the Visual Analogue Scale for Anxiety between the control group and the VR group for both women and their partners.

[e]VR: virtual reality.

The following variables showed a significant relation with ΔVAS-A: baseline VAS-A ($F_{1,75}$=8.4; P=.01) and history of CD ($F_{1,75}$=6.0; P=.02). These variables were incorporated in our multivariate regression, which significantly explained 16% of the variance in change in VAS-A score between time points 1 and 2.

As we used stratified randomization, we were able to perform a separate analysis based on the history of CD using an interaction term. Increase in the baseline VAS-A score at time point 2 (at admission) in women in the VR group with a history of emergency CD was 1.7 cm smaller than that in women with a history of emergency CD in the control group, although this effect was not significant (P=.06).

As there was a baseline difference in the level of education between both groups, we performed a regression analysis to analyze the effect of the different groups on the difference in VAS-A score, based on their level of education. By using dummy variables with university/college as a reference group, we were able to perform regression analysis on the categorical variables. If secondary education was the reference group, women had slightly lower VAS-A scores (0.2 cm). When vocational training or prevocational education was the reference group, the VAS-A scores were higher (0.5 and 0.2 cm, respectively). However, these results were not significant.

Secondary Outcome Measures

Median scores on the SSQ for motion sickness symptoms ranged from 1.0 to 1.5, reflecting the absence of discomfort caused by the VR video.

There was no significant difference in scores on the Tilburg Pregnancy Distress Scale subscales for both time points 1 and 2 between the VR group and the control group. The negative affect subscale showed higher scores at time point 2 than at time point 1 (control group, 25.2 [SD 3.9] vs 9.4 [SD 4.9]; VR group, 23.8 [SD 4.4] vs 9.1 [SD 4.4]), but this increase was equally present in both study arms.

For the PCQ questionnaire after delivery, we found a significantly higher score for the VR group without a history of emergency CD, indicating that they perceived a higher quality of care than the control group (10.2 [SD 3.8] vs 12.9 [SD 3.5]; P=.02). There was no significant difference between the control group and VR group of women with no history of emergency CD.

We received 26 completed questionnaires from time point 3 from women in which the question regarding the additional value of the VR video was filled in. Of the 26 women, 4 (15%) responded that they did not feel more prepared after seeing the VR video. The remaining 22 (85%) women responded positively. From the partners, 24 questionnaires were completed from time point 3 including this question, of which 19 (79%) partners responded positively. The remaining 5 (21%) partners did not feel more prepared after seeing the VR video.

Discussion

Principal Findings

In this study, we aimed to assess the effect of adding a VR video as part of standard preoperative information before planned CD

on the information level of women and their partners and the woman's level of anxiety. Our data showed that VR is not related to a significant decrease in self-reported preoperative anxiety for both women and their partners compared with a no VR condition. Only women with a history of emergency CD showed a trend toward decreased preoperative anxiety. With regard to the quality of care, women and their partners perceived a higher quality of care after watching the VR video.

Comparison With Previous Studies

Our preoperative VAS-A scores match those previously described in the literature [19,20]. As expected, anxiety levels were higher on the day of planned CD than at the baseline for both women and their partners, independent of the study arm. Differences in baseline characteristics of both groups, despite the randomized design of our study, might have influenced these results. Baseline anxiety level is an important predictor for preoperative anxiety, as stated in the literature and confirmed in our data. Baseline anxiety scores were slightly higher in the VR group compared with the control group for women in our study group, although this difference was not significant. In addition, there was a difference in the level of education between our VR and control groups at baseline, with the latter containing more women with a higher level of education. A higher level of education has previously been shown to be associated with increased preoperative fear [21]. Baseline anxiety scores were (nonsignificantly) higher in the VR group, although we would have expected the opposite based on the level of education. We could not reproduce this effect in our data through regression analysis. These baseline differences, therefore, probably did not significantly influence our primary outcome measure.

Besides the sociodemographic features (level of education and marital status), other factors such as presence of previous psychiatric disorders are known to influence preoperative anxiety. Women with a history of anxiety are more likely to experience fear of childbirth than others [22-25]. We could not reproduce this effect in our results. However, the percentage of women with a history of anxiety or depression included was low in both groups.

Another known risk factor for fear of childbirth is previous emergency CD [6]. In the case of emergency CD, women are rushed to the operating room because of complications that arise during labor, which creates a stressful situation. In the group of women with a history of emergency CD, adding the VR video to the standard preoperative information led to a 1.7 cm smaller increase in ΔVAS-A, with a *P* value at the border of significance (*P*=.06). However, given our cutoff value of 2 cm, we consider this result as not clinically relevant. Previous studies also reported lower preoperative anxiety in case of prior exposure to surgery [20,21,26,27]. Unfortunately, we have only taken into account previous (unplanned) CD and did not look at the history of surgery, in general, as a risk factor. Future studies should consider assessing for this confounder.

Literature about the use of (virtual) information videos in patient counseling report conflicting results of their effect on preoperative anxiety [28-33]. Anxiety is a multifactorial phenomenon, making it impossible to account for each interpersonal difference. Differences in study design concerning correction for confounding risk factors may influence the results. During the last two decades, there has been an increase in interest in the psychological impact of preoperative anxiety, both in general surgery and obstetrics. Besides (virtual) information videos, other methods to decrease preoperative anxiety have been studied, such as acupressure, intraoperative music therapy, and various information platforms [34-36]. VR has a big advantage of providing a visual reality–based experience. It offers a sense of *having been there*. This might help patients adjust their expectations toward the operation [37]. Furthermore, VR videos can be made in different languages very easily, and videos, in general, may be more accessible than written information for people with low literacy. However, as VR is a multisensory experience, motion sickness can occur in users. These complaints are mostly short lived but can cause real discomfort for users [38]. No cases of motion sickness occurred during our study after watching the VR video.

Recommendations

Although there was no significant difference in the main outcome measure, we found a trend toward decreased preoperative anxiety in the subgroup of women with a history of emergency CD after watching the VR video. This indicates that this subgroup of patients may benefit from this method of preoperative information. Careful selection of subgroups is the next step before implementing this information medium as part of standard care. Offering the VR video without obligation as part of the preoperative information at the outpatient clinic could help in gaining insight into the target population. Through tracking which and how many patients watch the VR video, it may be possible to assess the characteristics of this patient population interested in the video. However, patient privacy regulations make it difficult to facilitate this. Anonymous feedback questions provided through the mobile app after watching the video may add in retrieving this information. In addition, in the group of patients who waive the possibility of watching the VR video, it is important to ask for their reasons, which can be used in optimizing the content of the VR video.

Strengths and Limitations

With the rise of this innovative modality, a guideline toward the setup of clinical studies concerning the use of VR has recently been published. The authors recommend a three-phase development and validation process to uniformize the development and validation of VR applications [39]. Although our study was conducted before these guidelines were published, we believe our study comprises elements from all three phases suggested in the guidelines: We involved patients in the development process of the VR video (VR 1 phase), we conducted a randomized controlled trial to study the effect of VR on preoperative anxiety (VR 3 phase), and we briefly asked participants about their experience with the VR video (VR 2 phase). For future studies, adherence to these guidelines is desirable to ensure uniformity within the VR science platform.

There are several limitations to this study. First, power calculations were made based on the primary outcome measure (preoperative VAS-A score). Therefore, this study may be underpowered to show significant differences in results from the questionnaires designed to measure psychological

functioning in the perinatal period. The only significant result was a higher score on the PCQ questionnaire for the VR group without a history of emergency CD, indicating that they perceived a higher quality of care than the control group. However, the absolute difference between both groups is small and therefore not clinically significant. The relatively small study population is a limitation for each of the subanalyses performed in our study.

Second, we did not keep track of the number of people who refused to participate in the study and their arguments for refusal. This may have led to a selection bias because patients who are not keen on watching the VR video may be more likely to refuse when asked to participate. In addition, we did not verify if all participants who were randomized to the VR group actually watched the entire VR video. Although most participants watched the VR video at the time of inclusion, some chose to watch the video at home in their own environment. Therefore, there is a chance that some of them did not see the video, and including these completed questionnaires in our analysis could have caused a bias.

Fourth, there was a high amount of missing data because of missing questionnaires. We distributed paper questionnaires to patients at the time of inclusion and placed duplicate questionnaires at the ward where patients were admitted for CD. The researchers actively pursued the questionnaires. Despite our efforts, the percentage of missing questionnaires remained high, especially for questionnaire form 3. In future studies, electronic questionnaires available through a mobile app on a smartphone could facilitate this process.

Conclusions

Our study did not show a decrease in preoperative anxiety after VR information provision for patients undergoing elective CD. There was a trend toward decreased preoperative anxiety in the subgroup of women with a history of emergency CD who watched the VR video. Further research for identifying the characteristics of subgroups of patients who would potentially benefit from VR information provision is necessary.

Conflicts of Interest

SvR was the founder of Infor-Med BV. All other authors have no conflicts of interest to declare.

This randomized study was only retrospectively registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Translated version of the postdelivery questionnaire for partners (questionnaire 3).
[DOCX File , 13 KB - mental_v6i12e15872_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3489 KB - mental_v6i12e15872_app2.pdf]

References

1. Betrán AP, Ye J, Moller A, Zhang J, Gülmezoglu AM, Torloni MR. The increasing trend in Caesarean section rates: global, regional and national estimates: 1990-2014. PLoS One 2016;11(2):e0148343 [FREE Full text] [doi: 10.1371/journal.pone.0148343] [Medline: 26849801]
2. Perined. 2019. Healthcare yearbooks in the Netherlands 2017. URL: https://www.perined.nl/producten/publicaties/jaarboeken [accessed 2019-06-05]
3. Clement S. Psychological Perspectives On Pregnancy And Childbirth. First Edition. London: Churchill Livingstone; 1998.
4. Elsaharty A, McConachie I. Skin to skin: A modern approach to Caesarean delivery. J Obstet Anaesth Crit Care 2017;7(1):13. [doi: 10.4103/joacc.joacc_4_17]
5. Räisänen S, Lehto SM, Nielsen HS, Gissler M, Kramer MR, Heinonen S. Fear of childbirth predicts postpartum depression: a population-based analysis of 511 422 singleton births in Finland. BMJ Open 2013 Nov 28;3(11):e004047 [FREE Full text] [doi: 10.1136/bmjopen-2013-004047] [Medline: 24293208]
6. 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 Birth 2012 Sep;25(3):114-121. [doi: 10.1016/j.wombi.2011.06.001] [Medline: 21764400]
7. Storksen HT, Eberhard-Gran M, Garthus-Niegel S, Eskild A. Fear of childbirth; the relation to anxiety and depression. Acta Obstet Gynecol Scand 2012 Feb;91(2):237-242. [doi: 10.1111/j.1600-0412.2011.01323.x] [Medline: 22085403]

XSL•FO
RenderX

8. Truijens SE, Pommer AM, Heimel PJ, Verhoeven CJ, Oei SG, Pop VJ. Development of the Pregnancy and Childbirth Questionnaire (PCQ): evaluating quality of care as perceived by women who recently gave birth. Eur J Obstet Gynecol Reprod Biol 2014 Mar;174:35-40. [doi: 10.1016/j.ejogrb.2013.11.019] [Medline: 24332094]

9. Saisto T, Halmesmäki E. Fear of childbirth: a neglected dilemma. Acta Obstet Gynecol Scand 2003 Mar;82(3):201-208. [Medline: 12694113]

10. Szeverényi P, Póka R, Hetey M, Török Z. Contents of childbirth-related fear among couples wishing the partner's presence at delivery. J Psychosom Obstet Gynaecol 1998 Mar;19(1):38-43. [doi: 10.3109/01674829809044219] [Medline: 9575467]

11. Hobson JA, Slade P, Wrench IJ, Power L. Preoperative anxiety and postoperative satisfaction in women undergoing elective caesarean section. Int J Obstet Anesth 2006 Jan;15(1):18-23. [doi: 10.1016/j.ijoa.2005.05.008] [Medline: 16256338]

12. Kinnersley P, Phillips K, Savage K, Kelly MJ, Farrell E, Morgan B, et al. Interventions to promote informed consent for patients undergoing surgical and other invasive healthcare procedures. Cochrane Database Syst Rev 2013 Jul 6(7):CD009445. [doi: 10.1002/14651858.CD009445.pub2] [Medline: 23832767]

13. Infor-Med. URL: https://infor-med.nl/bibliotheek [accessed 2019-10-15]

14. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]

15. Facco E, Stellini E, Bacci C, Manani G, Pavan C, Cavallin F, et al. Validation of visual analogue scale for anxiety (VAS-A) in preanesthesia evaluation. Minerva Anestesiol 2013 Dec;79(12):1389-1395. [Medline: 23860442]

16. Pop VJ, Pommer AM, Pop-Purceleanu M, Wijnen HA, Bergink V, Pouwer F. Development of the Tilburg Pregnancy Distress Scale: the TPDS. BMC Pregnancy Childbirth 2011 Oct 26;11:80 [FREE Full text] [doi: 10.1186/1471-2393-11-80] [Medline: 22029691]

17. Truijens SE, Wijnen HA, Pommer AM, Oei SG, Pop VJ. Development of the Childbirth Perception Scale (CPS): perception of delivery and the first postpartum week. Arch Womens Ment Health 2014 Oct;17(5):411-421. [doi: 10.1007/s00737-014-0420-0] [Medline: 24663684]

18. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]

19. Hepp P, Hagenbeck C, Burghardt B, Jaeger B, Wolf OT, Fehm T, MAGIC Group. Measuring the course of anxiety in women giving birth by Caesarean section: a prospective study. BMC Pregnancy Childbirth 2016 May 18;16:113 [FREE Full text] [doi: 10.1186/s12884-016-0906-z] [Medline: 27188222]

20. Bansal T, Joon A. Anaesthesia, Pain & Intensive Care. 2017. A comparative study to assess preoperative anxiety in obstetric patients undergoing elective or emergency cesarean section. URL: https://www.apicareonline.com/a-comparative-study-to-assess-preoperative-anxiety-in-obstetric-patients-undergoing-elective-or-emergency-cesarean-section/ [accessed 2019-06-05]

21. Caumo W, Schmidt AP, Schneider CN, Bergmann J, Iwamoto CW, Bandeira D, et al. Risk factors for preoperative anxiety in adults. Acta Anaesthesiol Scand 2001 Mar;45(3):298-307. [doi: 10.1034/j.1399-6576.2001.045003298.x] [Medline: 11207465]

22. Saisto T, Salmela-Aro K, Nurmi JE, Halmesmäki E. Psychosocial characteristics of women and their partners fearing vaginal childbirth. Br J Obstet Gynaecol 2001 May;108(5):492-498 [FREE Full text] [doi: 10.1111/j.1471-0528.2001.00122.x] [Medline: 11368135]

23. Laursen M, Hedegaard M, Johansen C, Danish National Birth Cohort. Fear of childbirth: predictors and temporal changes among nulliparous women in the Danish National Birth Cohort. Br J Obstet Gynaecol 2008 Feb;115(3):354-360 [FREE Full text] [doi: 10.1111/j.1471-0528.2007.01583.x] [Medline: 18190372]

24. 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]

25. Spice K, Jones SL, Hadjistavropoulos HD, Kowalyk K, Stewart SH. Prenatal fear of childbirth and anxiety sensitivity. J Psychosom Obstet Gynaecol 2009 Sep;30(3):168-174. [doi: 10.1080/01674820902950538] [Medline: 19591052]

26. Moerman N, van Dam FS, Muller MJ, Oosting H. The Amsterdam Preoperative Anxiety and Information Scale (APAIS). Anesth Analg 1996 Mar;82(3):445-451. [doi: 10.1097/00000539-199603000-00002] [Medline: 8623940]

27. Kindler CH, Harms C, Amsler F, Ihde-Scholl T, Scheidegger D. The visual analog scale allows effective measurement of preoperative anxiety and detection of patients' anesthetic concerns. Anesth Analg 2000 Mar;90(3):706-712. [doi: 10.1097/00000539-200003000-00036] [Medline: 10702461]

28. Salzwedel C, Petersen C, Blanc I, Koch U, Goetz AE, Schuster M. The effect of detailed, video-assisted anesthesia risk education on patient anxiety and the duration of the preanesthetic interview: a randomized controlled trial. Anesth Analg 2008 Jan;106(1):202-9, table of contents. [doi: 10.1213/01.ane.0000287665.96156.72] [Medline: 18165579]

29. Eley VA, Searles T, Donovan K, Walters E. Effect of an anaesthesia information video on preoperative maternal anxiety and postoperative satisfaction in elective caesarean section: a prospective randomised trial. Anaesth Intensive Care 2013 Nov;41(6):774-781 [FREE Full text] [doi: 10.1177/0310057X1304100613] [Medline: 24180719]

30. Crabtree TD, Puri V, Bell JM, Bontumasi N, Patterson GA, Kreisel D, et al. Outcomes and perception of lung surgery with implementation of a patient video education module: a prospective cohort study. J Am Coll Surg 2012 May;214(5):816-21.e2. [doi: 10.1016/j.jamcollsurg.2012.01.047] [Medline: 22464659]

31. Sørlie T, Busund R, Sexton J, Sexton H, Sørlie D. Video information combined with individualized information sessions: effects upon emotional well-being following coronary artery bypass surgery--A randomized trial. Patient Educ Couns 2007 Feb;65(2):180-188. [doi: 10.1016/j.pec.2006.07.006] [Medline: 16959466]

32. Ruffinengo C, Versino E, Renga G. Effectiveness of an informative video on reducing anxiety levels in patients undergoing elective coronarography: an RCT. Eur J Cardiovasc Nurs 2009 Mar;8(1):57-61. [doi: 10.1016/j.ejcnurse.2008.04.002] [Medline: 18502689]

33. Bytzer P, Lindeberg B. Impact of an information video before colonoscopy on patient satisfaction and anxiety - a randomized trial. Endoscopy 2007 Aug;39(8):710-714. [doi: 10.1055/s-2007-966718] [Medline: 17661246]

34. Alanazi AA. Reducing anxiety in preoperative patients: a systematic review. Br J Nurs 2014;23(7):387-393. [doi: 10.12968/bjon.2014.23.7.387] [Medline: 24732993]

35. Agarwal A, Ranjan R, Dhiraaj S, Lakra A, Kumar M, Singh U. Acupressure for prevention of pre-operative anxiety: a prospective, randomised, placebo controlled study. Anaesthesia 2005 Oct;60(10):978-981 [FREE Full text] [doi: 10.1111/j.1365-2044.2005.04332.x] [Medline: 16179042]

36. Chang S, Chen C. Effects of music therapy on women's physiologic measures, anxiety, and satisfaction during cesarean delivery. Res Nurs Health 2005 Dec;28(6):453-461. [doi: 10.1002/nur.20102] [Medline: 16287051]

37. Riva G, Mantovani F, Capideville CS, Preziosa A, Morganti F, Villani D, et al. Affective interactions using virtual reality: the link between presence and emotions. Cyberpsychol Behav 2007 Feb;10(1):45-56. [doi: 10.1089/cpb.2006.9993] [Medline: 17305448]

38. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Appl Ergon 2018 May;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]

39. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]

Abbreviations

CD: cesarean delivery
CPS: Childbirth Perception Scale
PCQ: Pregnancy and Childbirth Questionnaire
SSQ: Simulation Sickness Questionnaire
VAS-A: Visual Analogue Scale for Anxiety
VR: virtual reality
ΔVAS-A: change in score on the Visual Analogue Scale for Anxiety

Viewpoint

Moderated Online Social Therapy: Viewpoint on the Ethics and Design Principles of a Web-Based Therapy System

Simon D'Alfonso[1,2], PhD; Jessica Phillips[2], BA; Lee Valentine[2,3], BA, MSocWk; John Gleeson[4], PhD; Mario Alvarez-Jimenez[2,3], PhD

[1]School of Computing and Information Systems, University of Melbourne, Parkville, Australia

[2]Orygen, Parkville, Australia

[3]Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

[4]School of Behavioural and Health Sciences, Australian Catholic University, Fitzroy, Australia

Corresponding Author:
Simon D'Alfonso, PhD
School of Computing and Information Systems
University of Melbourne
Grattan Street
Parkville, 3010
Australia
Phone: 61 399669387
Email: dalfonso@unimelb.edu.au

Abstract

The modern omnipresence of social media and social networking sites (SNSs) brings with it a range of important research questions. One of these concerns the impact of SNS use on mental health and well-being, a question that has been pursued in depth by scholars in the psychological sciences and the field of human-computer interaction. Despite this attention, the design choices made in the development of SNSs and the notion of well-being employed to evaluate such systems require further scrutiny. In this viewpoint paper, we examine the strategic design choices made in our development of an enclosed SNS for young people experiencing mental ill-health in terms of ethical and persuasive design and in terms of how it fosters well-being. In doing so, we critique the understanding of well-being that is used in much of the existing literature to make claims about the impact of a given technology on well-being. We also demonstrate how the holistic concept of eudaimonic well-being and ethical design of SNSs can complement one another.

KEYWORDS

Introduction

The modern omnipresence of social media and social networking sites (SNSs) brings with it a range of important research questions. One of these concerns the impact of SNS use on psychological well-being, a question that has been pursued in depth by scholars in the psychological sciences and the field of human-computer interaction (HCI) [1-5]. Despite growing multidisciplinary academic attention being devoted to questions concerning, for example, the "mental health outcomes associated with Facebook use" [6] over the course of the last decade, a related question concerning the design of social networking platforms remains underexamined, that is, *how do the design choices made by SNS programmers and related specialists contribute to the enhancement or the deterioration of a person's psychological well-being?* It is the goal of this paper to discuss not only certain strategic design choices in our development of an enclosed social networking platform for young people experiencing mental ill-health but also to critique the understanding of well-being that is used in much of the existing literature to make claims about the impact of a given technology on well-being. We will argue that the conception of well-being used to evaluate the effects of a given technology is oftentimes limited and inconsistent and may impair rather than enhance our understanding of how SNSs impact individuals. We will demonstrate how the holistic concept of eudaimonic well-being and ethical design of SNSs can complement one another.

As researchers who study persuasive technologies and design SNSs for mental health, we recognize that we can contribute in playing "a watchdog role for the human computer interaction

community in particular and the broader community" (p 230) [7] of consumers in general. This role includes the following 4 categories of action, as proposed by the persuasive technology pioneer Fogg [7]:

1. Identify artifacts and techniques: identify persuasive technologies and the persuasive techniques a researcher uses.
2. Examine effectiveness and effects: assess the effectiveness (intended impact) and the effects (unintended side effects) of persuasive technologies or strategies.
3. Disclose findings: those who study persuasive technologies have an obligation to disclose their findings.
4. If necessary, take or advocate social action: if a computer or artifact is deemed harmful or ethically questionable, a researcher should either take social action or advocate that others do so.

We will begin this paper by defining the term *social network* before proceeding with an overview of persuasive technology design and some associated ethical issues. Next, we will discuss the tendency in the existing literature to evaluate the effects of a given technology in relation to a limited concept of well-being that aligns with what researchers term *subjective well-being* (SWB) or *evaluation well-being* [8]. We will argue that the concept of SWB tells us very little about "what constitutes a well-lived life" [9] and, in turn, provides us with a narrow understanding of how SNSs impact individuals. Huang emphasized these difficulties in a 2010 meta-analysis, observing across 40 studies that well-being was inconsistently represented "by numerous psychological constructs, such as loneliness, depression, self-esteem, and life satisfaction," making it very difficult to draw any conclusions across the studies and contributing to mixed findings regarding the topic [10]. In this section, we will contrast *SWB* with what Ryff terms "a eudaimonic approach to psychological wellbeing" [9] and argue that the depth and breadth of this concept make it a valuable framework to evaluate the effects of an SNS platform and to guide the design and development of SNSs in youth mental health in particular. From here, we will provide an overview of the Moderated Online Social Therapy (MOST) platform we have developed and analyze it in terms of ethical persuasive design and eudaimonic well-being before outlining some new developments and discussing them in terms of the themes of this paper. Later topics covered include how MOST promotes a balance in online/offline activity and the incorporation of gamification.

The aims of this paper are in line with Fogg's categories of action as outlined above, focusing primarily on the first 2, that is, identifying artifacts and techniques that may render our technology persuasive and examining their effects and effectiveness, which involves assessing both the intended and unintended effects of our strategies. Disclosing our findings and advocating social action are, incidentally, actions performed in the writing and dissemination of this paper.

What Is a Social Network Site?

Scholars Boyd and Ellison differentiate between a social network site and a SNS. They define a social network site as [11]:

> *A web based service that allows people to 1) construct a public or semi-public profile within a bounded system; 2) articulate a list of others users with whom they share a connection; 3) view and traverse their list of connections and those made by others within the bounded system.*

They prefer the term social *network* site because these Web-based platforms are primarily designed to "enable users to articulate and make visible their [existing] social networks" (p 211) [11]. Put differently, they are not about making new friends or networks but about tending to their existing ones.

Networking, on the contrary, "emphasizes relationship initiation, often between strangers" (p 211) [11]. It is in this sense, however, that MOST is an SNS. Thus, adapting from Boyd and Ellison's definition, MOST is an SNS best conceived as a Web-based service that allows people to construct a public or semipublic profile within a bounded system and to initiate relationships generally between strangers. The vast majority of users who join the MOST SNS platforms for mental health do not know each other before use. Therefore, one of the primary goals of the interventions is for users to connect with others who share a similar lived experience but do not necessarily share a similar or the same offline social networks.

Persuasive Technology Design

The study of computers as persuasive technologies has roots in Fogg's seminal work and his coinage of the term *captology* from the acronym Computers as Persuasive Technologies [7,12]. According to Fogg, "a persuasive computing technology is a computing system, device, or application intentionally designed to change a person's attitudes or behavior in a predetermined way" [13]. Work in the field of persuasive technology design is particularly useful as a foundation for the development of digital technologies for behavior change and psychological well-being (PWB). Informed by Fogg's conceptualization of persuasive technology, Oinas-Kukkonen and Harjumaa [14] have developed a more concrete framework that transforms persuasive design principles into software requirements and system features. According to their Persuasive Systems Design model, there are 4 categories for persuasive system principles:

1. Primary task: The design principles in this category support the carrying out of the user's primary task and consist of reduction, tunneling, tailoring, personalization, self-monitoring, simulation, and rehearsal.
2. Dialog: The design principles in this category are about the feedback an interactive system provides to its users to help them move toward their goal or a target behavior. This category consists of praise, rewards, reminders, suggestion, similarity, liking, and social role.
3. System credibility: The design principles in this category describe how to design a system so that it is more credible

and thus more persuasive. The category consists of trustworthiness, expertise, surface credibility, real-world feel, authority, third-party endorsements, and verifiability.

4. Social support: The design principles in this category describe how to design the system so that it motivates users by leveraging social influence. The category consists of social facilitation, social comparison, normative influence, social learning, cooperation, competition, and recognition.

In his earliest expositions of captology, Fogg brought attention to the fact that "adopting an ethical perspective in this domain is vital because the topic of computers and the topic of persuasion both raise important issues about ethics and values" [7]. In the environment of the World Wide Web, social media, and ubiquitous personal devices such as smartphones, this topical theme takes on pressing significance.

Motivated primarily by commercial and advertising interests, many websites and apps incorporate features that are intentionally designed to hook users in, maximizing their attention and usage time without due regard for the quality or benefits to well-being of this usage. A movement has recently emerged to counter this phenomenon, colloquially termed *brain hacking*, by promoting the development of technology that is ethical and humane rather than addictive or of little genuine benefit. Founded by Tristan Harris, a former student of Fogg, the Center for Humane Technology (CHT) [15] is raising awareness of the issues associated with certain SNSs and smartphone technologies. The CHT writes that [16]:

> *Facebook, Twitter, Instagram, and Google have produced amazing products that have benefited the world enormously. But these companies are also caught in a zero-sum race for our finite attention, which they need to make money. Constantly forced to outperform their competitors, they must use increasingly persuasive techniques to keep us glued. They point AI-driven news feeds, content, and notifications at our minds, continually learning how to hook us more deeply—from our own behavior.*

The CHT is "creating humane design standards, policy, and business models that more deeply align with our humanity and how we want to live" [16]. The establishment of app design guidelines offers a simple and effective way to influence app architects by positively shaping the design and development of their systems. For example, the following is a selection of pertinent principles taken from a list published by the CHT that have been applied to the MOST platform [17]:

1. Does your product honor both on- and off-screen possibilities?
2. Does your product enhance relationships or keep people isolated?
3. Does your product land specific, net positive benefits in people's lives?
4. Does your product eliminate detours and distractions?

In our subsequent discussion of the MOST system, we will examine how it adheres to or embodies such persuasive design principles. Let us now discuss how well-being is traditionally conceptualized in the existing literature and provide the rationale

for why we have chosen to evaluate the MOST system in terms of a eudaimonic conception of well-being.

What Are We Talking About When We Talk About Well-Being?

In a systematic review on the "Impact of the Use of Social Network Sites on Users' Psychological Wellbeing," Erafni and Abedin [8] argue that research on *psychological well-being* has historically been aligned with 3 perspectives: the hedonic view, the eudaimonic view, and life satisfaction (also known as the evaluation well-being view). The hedonic and evaluation well-being views are frequently referred to as PWB in the literature and are rarely differentiated as different *kinds* of well-being. However, these distinctions, which we will discuss in a moment, are nontrivial, for when several measures of well-being are grouped beneath the broad umbrella term PWB, it becomes difficult to understand precisely *what* is impacted by a given technology.

SWB [18] concerns "moods and emotions [...] together labelled *affect* [and] represents people's on-line evaluations of the events that occur in their lives" (p 277) [19]. SWB consists of 3 components: "life satisfaction, the presence of positive mood, and the absence of negative mood, together often summarized as happiness" (p 144) [20]. Ryan and Deci argue that aligning well-being with happiness has a long history, dating back to the Greek philosopher Aristippus in the fourth century Before Christ. Thomas Hobbes, DeSade, and Jeremy Bentham followed in Aristippus's thinking and later enlarged this early philosophical hedonism [20]. The prevailing view present among hedonic psychologists today, Ryan and Deci argue, is the idea that well-being "consists of subjective happiness […] the experience of pleasure versus displeasure broadly construed [and] all judgements about the good/bad elements of life" (p 144) [20]. To illustrate how prevalent the concept of SWB is in discussions of SNSs, Erafni's and Abedin's analysis indicates that of 22 studies that met their criteria for inclusion, 15 used the measure of life satisfaction, 3 used the measure of affect or happiness, and 1 used the components we normally associate with eudaimonic well-being: autonomy, personal mastery, personal growth, positive relations, purpose in life, and self-acceptance [8]. However, it has been debated whether the SWB construct offers an adequate evaluation of a person's psychological wellness [21,22].

Carol Ryff has argued that prevailing concepts of SWB must be challenged as the construct's narrow focus on "assessments of feeling good, contentment and life-satisfaction" (p 13) [21] neglects "aspects of positive functioning such as purposeful engagement in life, realization of personal talents and capacities, and enlightened self-knowledge" [22]. Self-Determination Theory (SDT) pioneers Ryan and Deci have also challenged the SWB construct of well-being through their adoption of the concept of eudaimonia in their broader consideration of well-being and "what it means to actualize the self and how that can be accomplished" (p 146) [20]. SDT recognizes that 3 fundamental psychological needs are essential for the fulfillment of psychological growth. These include the need for autonomy, competence, and relatedness. For SDT, the fulfillment of these

3 basic needs is both the "natural aim of human life" and "typically fosters SWB as well as eudaimonic wellbeing" (p 147) [20]. In fact, a recent study on designing digital systems for motivation, engagement, and well-being shows how satisfying the 3 basic needs of SDT can increase these desired outcomes of user experience [23].

Ryff argues that SWB surfaced because of a mistranslation of the Aristotelian concept of *eudaimonia.* She writes that *eudaimonia*, when translated by Bradburn and other utilitarian philosophers of the 19th century, was taken to mean happiness [21]. Ryff argues that the trouble with this assumption is that it equates hedonia with eudaimonia, "something that was deeply contrary to Aristotle's distinction between the satisfaction of right and wrong desires" [21]. Furthermore, this conflation leaves out the essence of Aristotle's eudaimonia: "the striving towards excellence based on one's unique potential" [21]. Ryan and Deci too argue that Aristotle thought hedonic happiness and the pursuit thereof to be a *vulgar idea.* In their view, Aristotle believed that *true happiness* was to be found in "the expression of virtue" or "in doing what is worth doing" [20]. Therefore, Ryff's project has been to articulate a conception of PWB that is informed by and aligned with what she perceives to be the original essence of Aristotle's *eudaimonia.* Ryff's concept of PWB, the construct on which we will base our assessment of the MOST digital intervention for mental health, is concerned with 6 core components: self-acceptance, autonomy, personal growth, positive relationships, environmental mastery, and purpose in life [21].

It is evident then that SWB and PWB, although both aimed at understanding the greater question of what makes a good life, are underpinned by radically different value judgments about what that is. From the hedonic viewpoint of SWB, well-being is "equated with happiness and is formally defined as more positive affect, less negative affect and greater life satisfaction" [20]. PWB or eudaimonic well-being, as characterized by Ryff's construct, in contrast, conceives of well-being as a broad and complex concept consisting of 6 intersecting variables (as above) and has little to do with the pursuit of pleasure and a diminishment of negative affect. PWB recognizes that the pursuit of well-being is frequently characterized by periods of negative affect, especially when one pursues a meaningful and difficult goal.

What this discussion of SWB and PWB reveals is that the construct of well-being is *controversial and unresolved* [20]. For that reason, it is less than straightforward to interpret claims made about how technology impacts individual well-being. Furthermore, the overreliance of SWB measures to assess the impact of technology suggests that we are ascertaining a very narrow understanding of technology's impact. An additional point worth emphasizing is that psychology as a discipline has for the last century been concerned with "the amelioration of psychopathology" and not with the enhancement of well-being and individual growth [20]. Therefore, PWB is central in our discussion of the nexus between mental health and SNSs because the focus for psychology as a discipline has shifted, and we are now witnessing an increased focus on recovery in the fuller sense, including social and economic participation. We want to now bring these focal changes to bear on SNS design in

particular. In the section to follow, we turn to a discussion of the MOST system in terms of ethical persuasive design and well-being considerations, showing how MOST fosters this broader sense of PWB or eudaimonic well-being.

Moderated Online Social Therapy

The MOST project, based at eOrygen (the digital mental health division of Australia's Orygen youth mental health centre), has been researching and developing online social therapy systems for mental health since 2010. Consisting of a multidisciplinary team of clinical psychologists, computing and information systems researchers, software developers, creative writers, illustrators, and peer workers, MOST has been primarily concerned with evaluating the efficacy of online therapy and developing engaging digital technologies for young people experiencing mental ill-health. Trials have also been adapted to nonyouth cohorts, specifically the families/carers of such young people.

The result of this work has been an online social therapy framework powered by the MOST Web platform (or MOSTware), which integrates Facebook-style social networking, specialized therapy units, and a forum-like feature where users can pose and cooperatively crowdsource solutions to common problems [24-26]. All of this occurs within a clinical and peer-moderated environment.

As will be detailed later on, the MOST framework has been inspired by the field of positive psychology. Thus, on this basis, it technologically embodies an approach that balances a traditional focus on psychopathology with an emphasis on positive human development and flourishing [27,28]. Together with a participatory design approach involving the users of our systems (young people and their carers) [29], "we became aware of evidence highlighting that a focus on deficits in online systems can lead to demoralization and disengagement and so we began to specifically draw upon positive psychology frameworks" [30]. More generally, MOST, as a technology that supports well-being and flourishing, aligns with the field of positive computing, which is itself informed by positive psychology and provides a foundational framework for the "design and development of technology to support psychological wellbeing and human potential" [31].

By tailoring therapy content to target the treatment of specific conditions and adding any required code customizations, the flexible MOST platform enables the setting up of individual sites for a variety of mental health cohorts. To date, MOSTware has powered several successful studies, including the following:

- Rebound, a pilot trial for relapse prevention of major depressive disorder in young people [32].
- Meridian, a pilot trial for carers of young people diagnosed with depression and anxiety [30].
- Momentum, a pilot trial for young people at ultrahigh risk for psychosis [33].
- Altitudes, a 2-year randomized controlled trial (RCT) of MOST for carers of young people with psychosis [34].
- Horyzons, a 5-year RCT of MOST for relapse prevention following a first episode of psychosis [35]. This trial will

be used throughout this paper for illustration of the MOST platform.

- Generation, a public trial for general help-seeking youth, in collaboration with the eheadspace [36] electronic mental health service [37].

System credibility (as described in the section Persuasive Technology Design) within these interventions is ensured by several factors and measures taken:

1. Therapy content is created by expert psychologists, young people with a lived experience of mental ill-health, writers, and illustrators.
2. System moderation is conducted by qualified clinical psychologists and youth workers.
3. Peer moderation is conducted by young people with a lived experience of mental ill-health who model system usage and interact with clients.
4. Clinical trials are conducted with established protocols and within a framework of formal ethics approval.

Social Network

In the social network newsfeed section of the system otherwise known as The Café (Figure 1), users can contribute posts and comments, share experiences and interests, give and obtain support, gain new perspectives, and search for job opportunities and information in the Job Zone. The newsfeed feature is such that all users see the same newsfeed (ie, same ordered list of posts), though new developments will see the introduction of an algorithmically driven newsfeed tailored to each individual user. In Team Up, a relatively new feature to be discussed in more depth later, users can set a challenge for themselves that others can also participate in or follow as members of a *cheer squad*. MOST also contains an instant messenger service for communication between users, though this is only activated in the social network section of the site so as not to distract users with message popups while they are engaging with other parts of the site, particularly therapy content that requires deep attention. This design choice aligns with the CHT principle concerning the elimination of product detours and distractions.

Figure 1. The Moderated Online Social Therapy news feed.
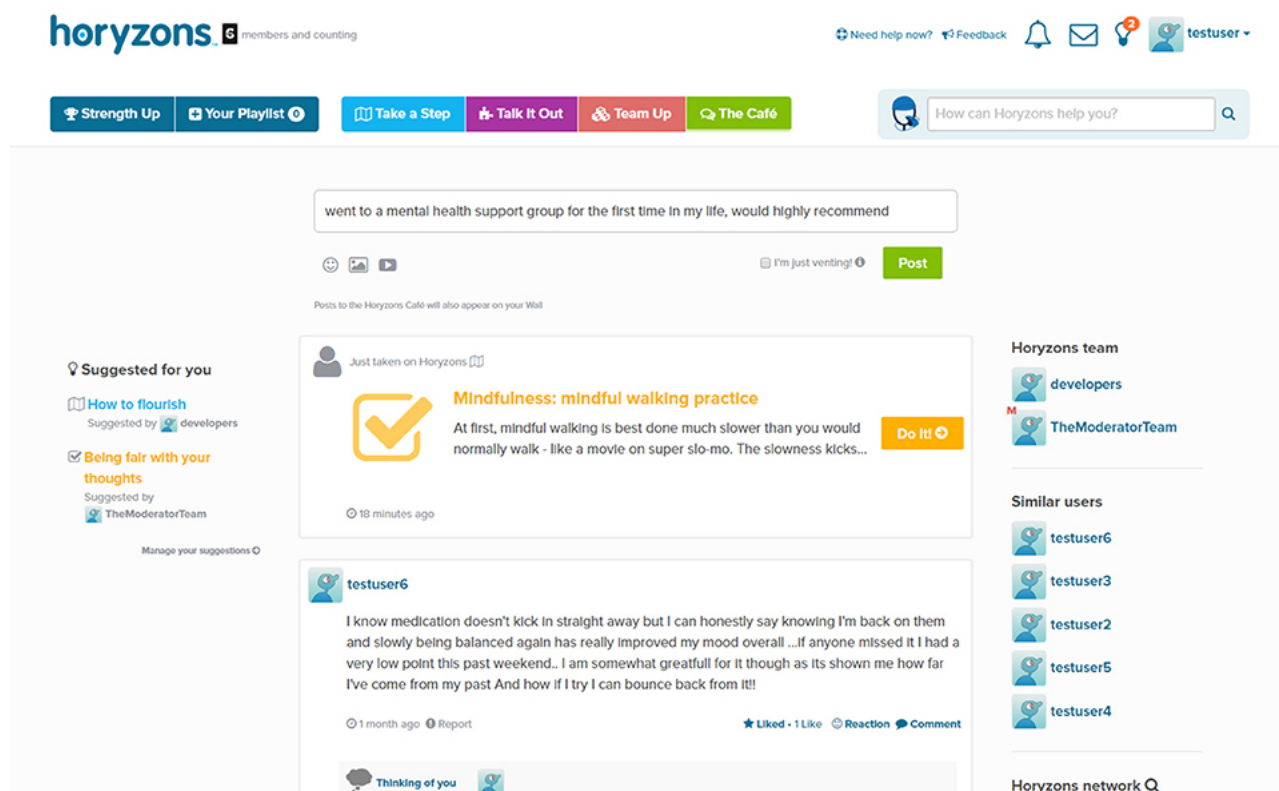
Steps and Actions

The 2 types of therapy units offered in the MOST system are *Steps* and *Actions*. Steps are psychoeducational therapy modules that assist people with developing psychological skills, such as self-compassion and mindfulness, and core components of eudaimonic well-being or PWB, such as purpose in life, autonomy, personal growth, environmental mastery, positive relationships, and self-acceptance. Although there is still a case to be made for helping users to regulate unnecessary and pointless suffering by reducing or moderating unhelpful and oftentimes distressing emotional states, the ultimate aim of Steps and Actions is to foster PWB, particularly in relation to autonomy and self-acceptance. These standalone therapy units can also be found and sequentially completed within themed Step collections called *Pathways*. Steps are delivered as engaging content, including low-literacy comics, and have been developed collaboratively by clinical psychologists, creative writers, young people, and graphic designers. In terms of dialog support [14], as introduced in the section Persuasive Technology Design, therapy content has been developed to imitate its users with relatable material and suitable language for the target cohort (similarity) with an appropriate look and feel (liking). Figure 2 illustrates an example Step, *How to flourish*, which is primarily
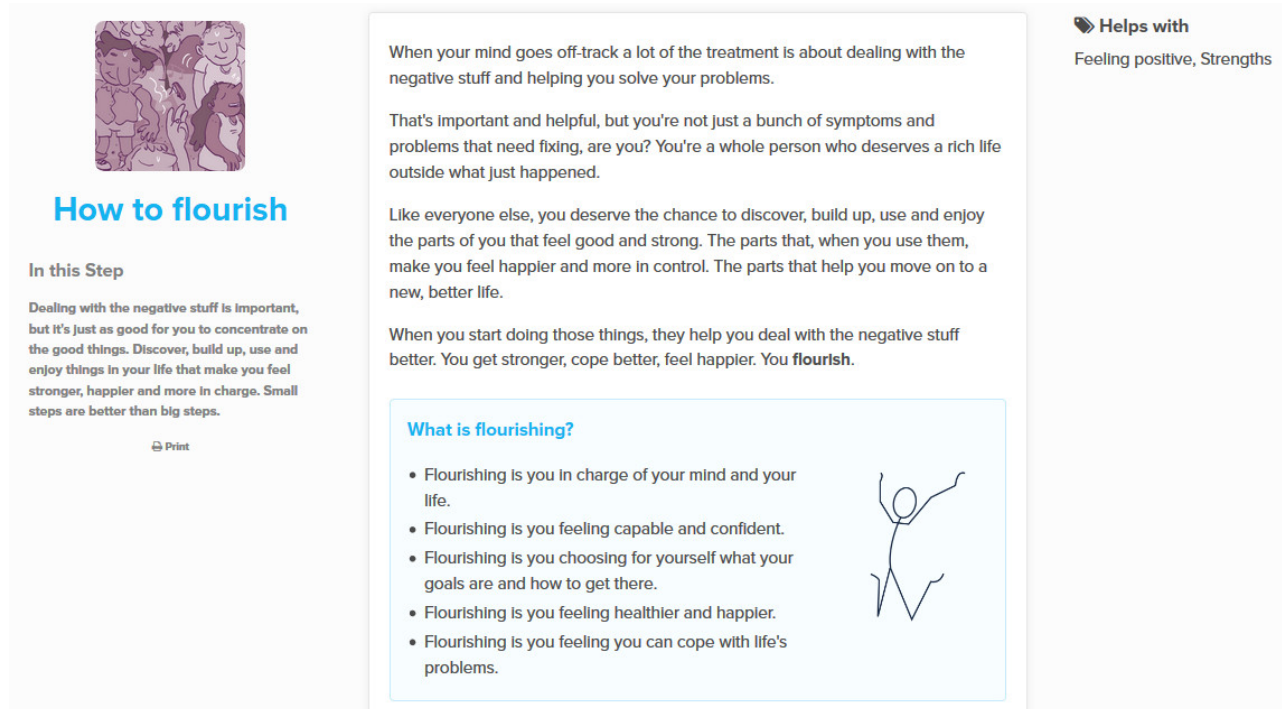
concerned with supporting young people to locate purposeful activities and meaning in their lives.

Figure 3 illustrates part of another example Step, *Everyday mindfulness*, in the newer comic format.

Social interaction is embedded within Steps through Talking Points, which are focused questions that invite young people to discuss and share their own experiences. Figure 4 shows a Talking Point within the Step *How to flourish*.

Figure 2. The Step How to flourish.

Figure 3. The Step Everyday Mindfulness.

Figure 4. A Talking Point for the Step How to flourish.
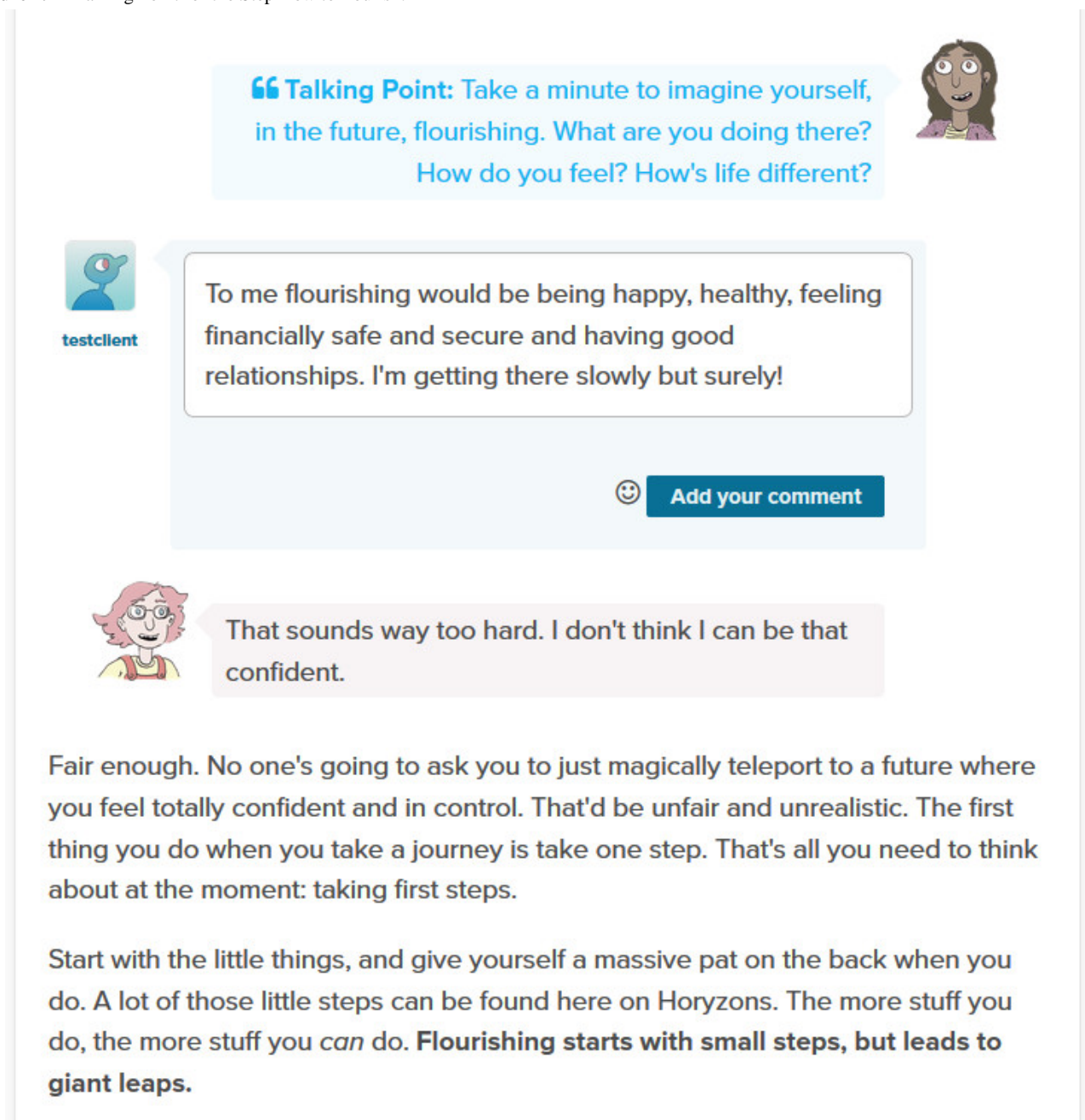
Actions are bite-sized behavioral activities or tasks that users can do to apply mindfulness, self-compassion, and personal strengths to relevant real-world situations. The use of context-specific, action-based suggestions through online interventions has been recommended to change behavior, develop skills, and increase practice and generalization of these skills to real-life situations [38]. Users are responsible for completing the action in their own time and are encouraged to report back to others with their progress. Through choosing relevant and personally meaningful actions, the intention is that users come to recognize that they have the capacity to shift things in their lives for the better. Put differently, actions aim to foster a sense of environmental mastery through autonomy by urging users to recognize and indeed experience the degree of control they have over many of the circumstances of their lives. Certain Actions are embedded within the Steps to which they are relevant, so that once a Step is completed, users can do an Action to reinforce what they have learned (Figure 5 shows an Action that is embedded in the Step *Everyday mindfulness*). Users can also engage with Actions through the Strengths to which they are connected; if an Action is connected to a Strength, then doing that Action will help to promote that Strength.

Figure 5. The Action Mindfulness: deepening your awareness practice.
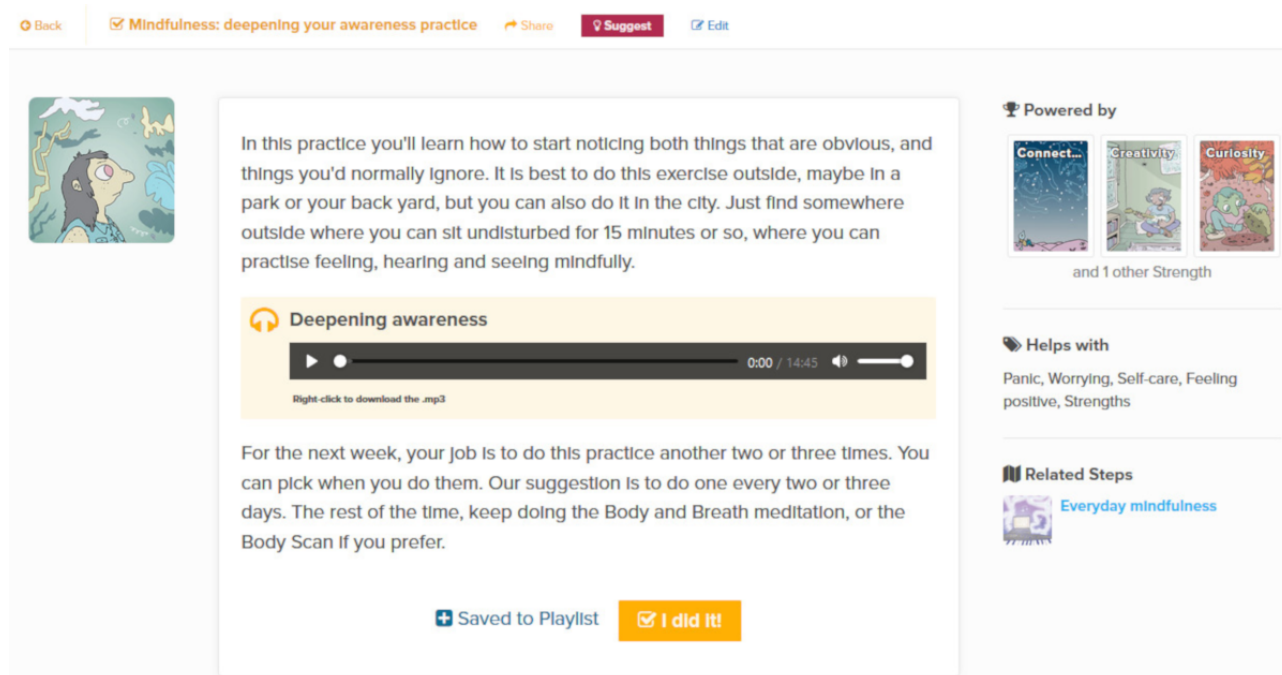
Strengths

Upon joining the system, users have the option to complete an initial exercise where they choose 5, out of 24, character strengths that they believe best apply to them. Although they are free to partake in activities associated with any of the Strengths, content associated with the Strengths a user selects will be promoted by the MOST system and used to guide the user through the experience over time. An illustration of the Strengths page after completion of this initial exercise is given in Figure 6.

Figure 6. The Strengths page.
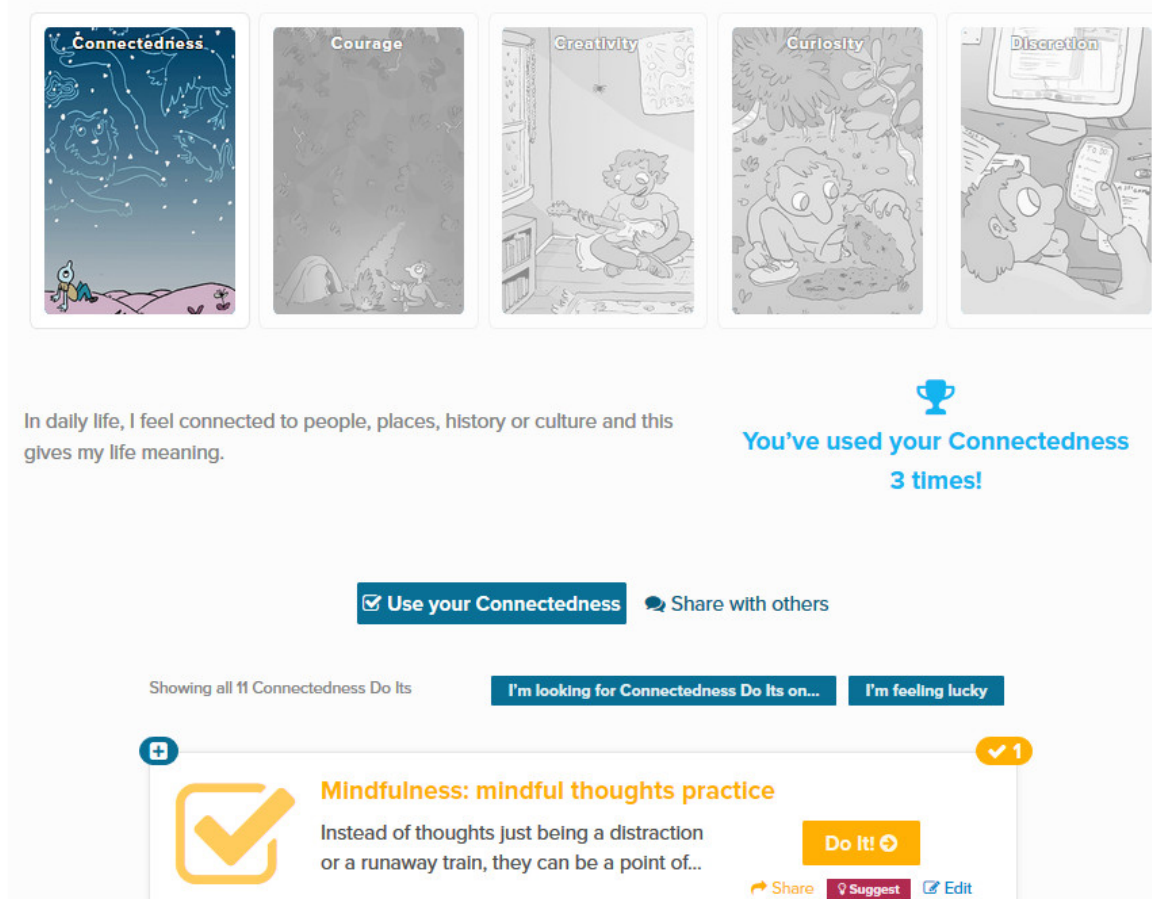
The set of Strengths available in the MOST platform is not fixed and can be customized according to the purpose of a particular intervention. The most representative set of Strengths, found in interventions such as Horyzons, consists of 24 strengths drawn from the field of positive psychology [39]. In addition to those illustrated in Figure 6, others include fairness, forgiveness, gratitude, and honesty.

Connected with the ideas and frameworks discussed in the earlier section on well-being, the positive psychology movement is defined as "the scientific study of positive human functioning and flourishing on multiple levels that include the biological, personal, relational, institutional, cultural, and global dimensions of life" [28]. A complement rather than replacement for traditional areas of psychology, positive psychology is concerned with human flourishing, a doctrine where the concept of eudaimonia, that is, "the striving towards excellence based on one's unique potential" [21], is deemed central to a life lived well. Positive psychology understands the good life as that in which one uses their "signature strengths every day to produce authentic happiness and abundant gratification" [40]. The 24 Strengths drawn from positive psychology and embedded in the MOST system correspond neatly with the building blocks of eudaimonic well-being. By encouraging users to identify strengths they currently possess and those they wish to build, the system strives to promote self-acceptance, self-competence, and personal growth. For instance, users come to recognize and identify their current positive attributes (self-acceptance). An invitation that may be infrequently offered in the offline world, users also come to value personal growth through the recognition that their personal circumstances are not fixed; rather they can indeed *grow* or develop strengths that they do not currently believe they possess.

Thus, the positive psychology–inspired Strengths and therapy features of MOST naturally align with eudaimonia and Ryff's concept of PWB. Although MOST interventions strive to maintain or improve levels of SWB, they also ultimately aim to promote eudaimonic or PWB.

Talk It Out

The Talk It Out (TIO) feature is a space where users can nominate problems or difficulties that they are seeking to discuss with other users. TIO is based on an evidence-based cognitive behavioral therapy social problem-solving framework developed by Nezu et al [41]. Once a user has nominated a problem and framed it together with a peer worker, they brainstorm solutions together with others and discuss the pros and cons of each suggestion. Those moderating the process then synthesize the content and *wrap it up*. Apart from the role each of these TIO discussions play at the time in providing an active forum to help a person navigate their way through a problem, the overall result is an invaluable user-generated *knowledge base* repository that can be searched and referred back to by users at any time.

Environmental mastery is perhaps the most pertinent aspect of eudaimonic well-being fostered by TIO. Through the various phases, participants learn how to approach problem solving in a *systematic and thoughtful* way [41]. They learn how to clearly and unemotionally define the problem they are facing, and in doing so, they are better positioned to ask for help and others

are more likely to provide help that is relevant to their particular problem. The second phase of the TIO, brainstorming, further contributes to a sense of environmental mastery as it helps participants to recognize that when they perceive that there are multiple ways to solve a well-defined problem, they become less stressed and feel more control over their lives, safer, and more hopeful about locating a solution [41]. Finally, brainstorming contributes to a sense of environmental mastery by inviting people to externalize their problem [41]. Thus, rather than ruminating on their own thoughts, brainstorming removes the isolation that may accompany an internalized problem and supports participants to recognize that by externalizing a problem to others (or a network of other people like them) they can feel more in control of locating a solution.

In line with the CHT guidelines listed in section Persuasive Technology Design, TIO has been developed to foster connectivity and to help counter isolation while also promoting a sense of autonomy. The TIO feature is nonhierarchical—there is no patient and there is no expert. Rather, all participants are invited to be the experts on their own lives. This setup aims to foster a sense of autonomy as it removes the need to defer to a clinician when seeking a solution to an everyday problem. Through the TIO, participants can come to recognize that their lived experience is valuable and can be used, whether they are in a formal role as a peer worker or not, to support others. Therefore, TIO should bring people together and empower them to both contribute to and seek social support.

Unlike other social media and networking sites, the MOST platform has been specifically developed for the online treatment of mental ill-health. Naturally, therefore, a paramount consideration is the clinical efficacy and HCI of its specialized system components, such as the TIO forums, Steps and Actions therapy units, and moderator involvement. However, beyond these technical and clinical considerations and in concert with them, we must be mindful of the following:

1. Ensure that the system in general is not detrimental to well-being and that features and components of little value or those that have a negative impact on psychological health are avoided.
2. Develop content and features that promote mental health and well-being and that users are motivated to use for good reasons.

In describing the MOST system throughout this section, we have covered how some of these considerations have been incorporated into its design. We now turn to discuss some new MOST developments as they relate to the themes of this paper.

Balance Between Online and Offline Activities

Commercial SNSs often aim to maximize their sphere, not only in terms of user attention and usage time but also in terms of bringing and containing as much activity as possible within the site. Contrary to such monopolization, we maintain the following:

1. A website should encourage and facilitate offline activity outside of the site.
2. Online activity should be promoted within the site insofar as it is something that is beneficial and inherent to the site or can genuinely only be mediated/facilitated by the site.

This approach ties in with the CHT design principle "does your product honor both on and off-screen possibilities" and helps to deter social media and networking addiction [42,43]. We posit that it also accords with the eudaimonic conceptualization of well-being in the sense that the good life cannot be one confined to the digital realm. The resulting design philosophy

is embodied in several existing and upcoming features and functionalities associated with the MOST system.

Team Up

The *Team Up* feature is a relatively recent addition to the MOST platform, and a minimum viable implementation has been tried out on a select few trials. Team Up begins with a user nominating a challenge or goal that they want to achieve by a certain date. Other users can join the challenge either as a fellow participant or a supportive follower. An example Team Up challenge, "I want to exercise twice per week," is captured in Figure 7. As can be seen, Team Up embodies several social support persuasive system principles [14]:

Figure 7. An example Team Up challenge.
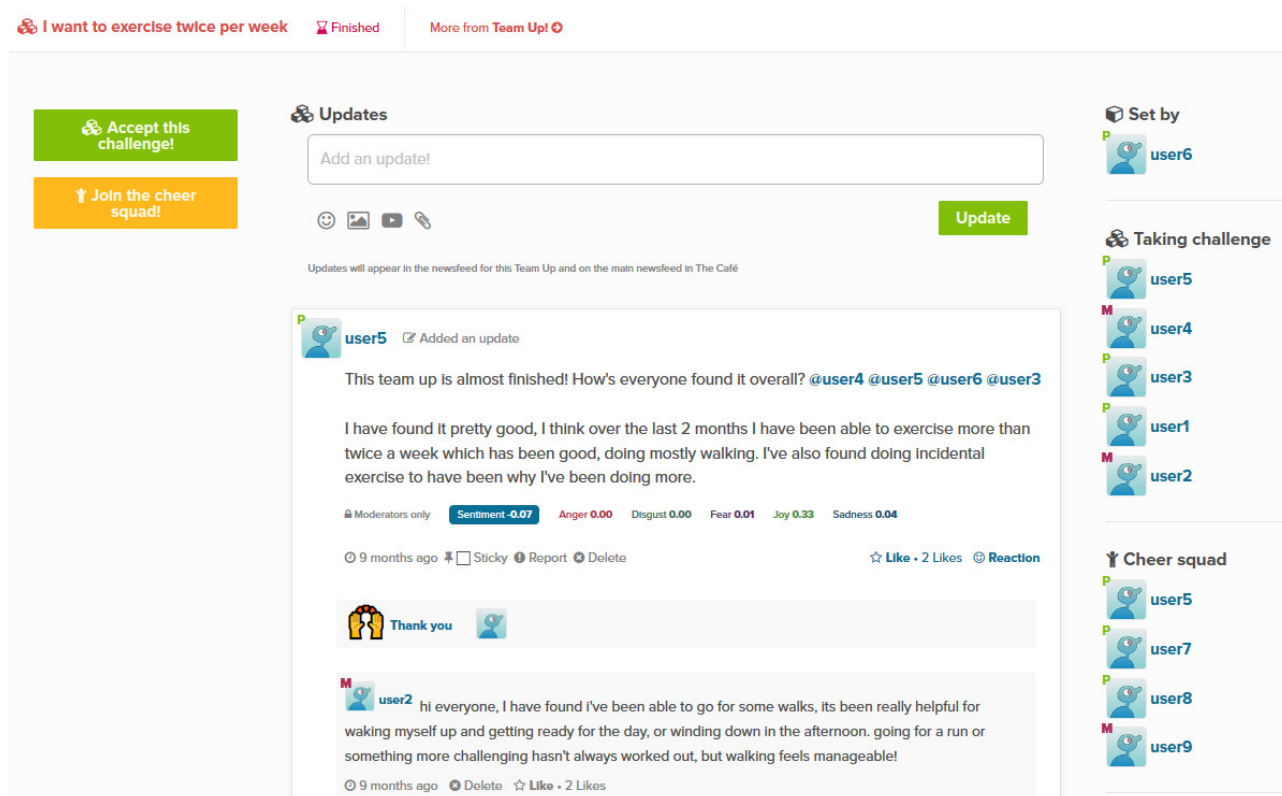
1. Social learning: A person will be more motivated to perform a target behavior if (s)he can use a system to observe others performing the behavior.
2. Social facilitation: System users are more likely to perform target behavior if they discern via the system that others are performing the behavior along with them.
3. Recognition: By offering public recognition for an individual or group, a system can increase the likelihood that a person/group will adopt a target behavior.

However, other social support design principles such as competition and normative influence or peer pressure seem inappropriate given the type of supportive atmosphere being fostered in MOST.

Team Up demonstrates a commitment to ensuring that the systems strike a balance between online and offline activity. This simple feature drives an activity loop whereby users initiate their intention within the MOST system, and they then proceed to carry out activity in the offline world. They then return online

to the MOST system to report their progress, and this cycle can continue. Crucially, without some activity in the offline world, there is no point in using the online feature, and thus, it encourages healthy activity offline, promotes social support and interaction within the online community, and uses the site only to facilitate these 2 things. Support and validation are offered to the Team Up challenger by other challengers and supporters joining, and this in turn promotes and encourages the completion of a beneficial goal. Interestingly, in a review of persuasive design in internet-based psychological therapy systems for adolescent depression [44], it is claimed that "persuasive design features that leverage social support to motivate users were rarely, if ever, reported features." Such opportunities should not be missed, and the Team Up feature is one such example.

Online and Offline Therapy

The complete therapy experience in MOST consists of both online and offline units. Steps, which can be considered the digital counterpart to traditional face-to-face therapy, are

engaged with and contained within the online MOST platform; users read the content and rate, share, and comment on screen.

However, Actions, which are often embedded within the Steps they are related to, generally involve doing an offline exercise to reinforce skills, especially those covered in the Step. Once a Step is completed by the user online, relevant Action suggestions populate the bottom of the Step screen, thus offering *choices that send the user off the screen*, as per the guidelines offered by the CHT.

The fact that Actions are embedded and confined within the MOST platform does, however, present a certain limitation to their utilization. This is because although access to the library of Actions is presently confined to the MOST Web app via standard access points such as lists and a search bar, they generally involve activity that is done offline and outside of the platform in relevant real-world scenarios. To remove these online to offline barriers and facilitate Action engagement, a new Action delivery system in the spirit of just-in-time adaptive interventions is in development [45]. The idea is to respectfully use smartphone sensors to determine information about a user's present psychological states and/or their current situation or location and then use this information to deliver to them, using push notifications contextually relevant and helpful Actions they can do in situ, in real time [46].

For example, the Action "Mindfulness: deepening your awareness practice" (Figure 5) contains an audio guide on mindfulness with the suggestion that it be done outside in a park. Mobile phone geolocation data could be used to determine if a user is spending time in a park and deliver such an Action suggestion. As another simple example, the detection of a bout of nocturnal phone activity that deviates from one's standard usage times or is at odds with their indicated chronotype could lead to the suggestion of an exercise for insomnia.

Given such a system, a user could gain therapy benefits without having to log into the MOST platform. The aim is for such a system to reduce entry barriers into therapy completion and increase motivation and the number of Actions a user fulfills. Thus, beyond simply offering and expecting users to visit a website for help, the system can meet users in real-life situations and dynamically deliver accessible and effective personalized help directly to their pockets. In terms of the persuasive system principles by Oinas-Kukkonen and Harjumaa [14], this mobile therapy delivery system embodies the primary task support principles of tailoring (information provided by the system will be more persuasive if it is tailored to the potential needs, interests, personality, usage context, or other factors relevant to a user group) and personalization (a system that offers personalized content or services has a greater capability for persuasion).

Such a detect-and-deliver system may also inhibit detrimental smartphone usage. In their article on 10 lessons learned about SNSs and addictions, Kuss and Griffiths (p 8) [43] list "Smartphone Addiction May Be Part of SNS Addiction" and write the following in their discussion of this item:

> *According to the pathway model, an addictive pattern of mobile phone use is characterized by the use of*

> *specific applications, including calls, instant messaging, and the use of social networks. This suggests that rather than being an addictive medium per se, mobile technologies including smartphones and tablets are media that enable the engagement in potentially addictive activities, including SNS use.*

In terms of this issue, the benefit of such a detect-and-deliver system is that the smartphone becomes a tool to deliver contextually relevant Actions suggestions in offline situations rather than a necessary portal to an online app.

Gamification

Gamification, as standardly defined, is the use of "video game elements in nongaming systems to improve user experience (UX) and user engagement" [47]. Not surprisingly, there is an overlap between gamification and persuasive technology [48]. Gamification ties in with the Dialog Support Rewards principle (systems that reward target behaviors may have great persuasive powers) [14], and "some persuasion mechanisms can be regarded as similar to those applied in gamification, such as feedback and rewards" [49]. There are also certain ethical dimensions of gamification to consider, including the argument that gamification is at odds with human flourishing and that it could be "morally corrosive by adversely impacting character" [50]. There is currently work being done on adopting gamification to promote (mental) health and well-being [51-54]. We see the potential to positively harness gamification to enhance the adoption and sustained use of persuasive technologies that promote positive behavior change for people, particularly youth, in the mental health sphere.

Although not a formative consideration during the initial design and development of the MOST system, developments have since started to incorporate some gamification to encourage social engagement and therapy participation. This is in part because of feedback from users who have expressed a desire for the incorporation of gamification, particularly as ascertained from recent posttrial interviews for a qualitative study in development. In designing gamified components, we recognize that it is important to ensure that gamification does not negatively impact users and is not exploited by other parties. When done appropriately (see the studies by Kim and Werbach [55] and Llagostera [48] for an overt example of inappropriate and pernicious gamification), competitiveness may be a suitable and effective ingredient in a gamified system. However, given that MOST is designed for young people experiencing a variety of mental health conditions, they may be particularly vulnerable or sensitive to potential negative effects of inciting competition and social comparison. Thus, the focus is on personal gamification and opportunities to make gamification a social, but not necessarily competitive, experience. In game theoretical parlance, we are dealing with non-zero-sum, possibly cooperative, games.

As introduced in the section Moderated Online Social Therapy upon joining the system, users can complete an initial exercise where they choose 5 out of 24 Strengths that they believe best apply to them. Changes to the Strengths system, at this stage in prototype form, involve adding a gamification component.

Rather than a user selecting a fixed group of their top-5 Strengths when they begin using the site, with this new system, each Strength starts out as being more or less equal and various bits of user activity in the site over time add to the points score that they achieve for each strength. For example, contributing a newsfeed post could add 1 point to a user's Social Connectedness strength score. As another example, users could gain Strength points in their Supportiveness strength by contributing a response to another user's TIO problem.

The TIO feature is particularly interesting in terms of gamification considerations. As described earlier, the basic flow of a TIO discussion starts with the suggestion of a problem for communal discussion. Once this problem has been shaped, brainstormed solutions are offered by participants. A list of established solutions is then discussed by the group, where for each solution offered, users can respond to the solution with commentary responses in the form of *pros* and *cons*. Unlike the MOST news feed, in which posts and comments can be responded to via likes and reactions, no such response options to user posts have been implemented in TIO. The forum-like, crowdsourced question-and-answer structure of TIO makes it amenable to a style of reputation system gamification, in a manner similar to that of highly popular crowdsourcing forums such as Stack Overflow, where users are incentivized to contribute responses to questions by a points-and-recognition system. Users earn reputation points for offering answers, having their answers upvoted or downvoted by peers, and can receive badges for their valued knowledge contributions [56-58].

Despite the suitability and tremendous success of Stack Overflow's gamified reputation system, the viability or ethicality of such a system in TIO is problematic given its sensitivities. To begin with, Stack Overflow has implemented a well-crafted downvoting option, for which there is a good informational value rationale as downvoting another user's contribution gives "you the critically important ability to distinguish between the good, the bad, and the ugly" and the ability to "tell the difference between a post that is harmless but uninteresting, and one that is actually wrong or harmful" [59]. However, TIO contributions are not pure informational entities in the way that *Stack Overflow* posts are. They are often responses inspired by lived experience and may be imbued with emotionality and personal sharing. Therefore, it is important to preclude negativity and the perceived hostility that may arise through downvoting. Furthermore, any possibility of categorizing right and wrong answers is problematic. For forums such as Stack Overflow that trade in questions requiring answers of a factual nature, this is more straightforward, though not always perfectly clear. For TIO, the nature of the problems raised, and responses given, is such that any conception of right and wrong is meaningless, highly ambiguous, or unfeasible. Although there is the possibility of implementing a voting system restricted to upvoting to promote positivity and incentivize contributions, a

resulting problem could be that an implicit competitive ranking emerges.

The gamified Strengths system differs in this regard as it offers a form of individual incentivization without the issues and pressure associated with social comparison and competition. It is also not intended to be an attempt at pointsification (cases where a stock approach of gamification has been added on top of an existing system [60]) simply to generate extrinsic motivation through the collection of awards. Rather, with a dash of gamification, the Strengths points system should serve to provide users with a way to track their activity levels on the site, the skills their activity is building, and the strengths they are actualizing. Hopefully, these are goals for which the user has an intrinsic motivation to use MOST.

Conclusions

Digital technologies, particularly computers, have been studied as persuasive technologies for at least a couple of decades now. Beyond the general aim of designing products that people want to use, there is the goal of creating interfaces and features designed to encourage certain actions or to change a person's attitudes and behaviors. However, it has become apparent that given the commercial pressures and the race for user attention, techniques of persuasion employed by pervasive social media and networking sites are more about capitalizing our attention and generating usage that is addictive rather than usage that prioritizes well-being and is in the user's best interest. A growing awareness of this phenomenon and the establishment of a movement to counter it have provided us with an impetus to reflect upon and analyze the MOST therapy framework and Web platform developed for online mental health interventions. We have discussed in this paper how consideration of these issues has shaped our development of the MOST system, with the intention to foster usage and engagement that is conducive rather than detrimental to mental health and well-being.

We have also had the opportunity in this paper to critically explore the notion of well-being. We contend that although hedonic or SWB is important, it is problematically limiting to confine well-being evaluations of SNSs to this type of well-being. Rather, a notion of well-being rooted in the Aristotelian conception of eudaimonia is another important dimension of well-being to consider. Although analyses of certain commercial SNSs have been restrictively conducted only in terms of SWB, given the positive psychological and therapeutic nature of the interventional MOST system, we are in a prime position to foster this sense of eudaimonic well-being.

The MOST framework is an evolving one, and as development continues on systems and features powered by tools, technologies and methodologies from artificial intelligence, ubiquitous computing, and HCI, we must remain mindful to scrutinize these developments in terms of ethical design so that user well-being remains paramount.

Acknowledgments

Conflicts of Interest

The authors are part of the MOST development team.

References

1. Chou HG, Edge N. 'They are happier and having better lives than I am': the impact of using Facebook on perceptions of others' lives. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):117-121. [doi: 10.1089/cyber.2011.0324] [Medline: 22165917]

2. Deters FG, Mehl MR. Does posting Facebook status updates increase or decrease loneliness? An online social networking experiment. Soc Psychol Personal Sci 2013 Sep 01;4(5). [doi: 10.1177/1948550612469233] [Medline: 24224070]

3. Kross E, Verduyn P, Demiralp E, Park J, Lee DS, Lin N, et al. Facebook use predicts declines in subjective well-being in young adults. PLoS One 2013;8(8):e69841 [FREE Full text] [doi: 10.1371/journal.pone.0069841] [Medline: 23967061]

4. Steers MN, Wickham RE, Acitelli LK. Seeing everyone else's highlight reels: how Facebook usage is linked to depressive symptoms. J Soc Clin Psychol 2014;33(8):701-731. [doi: 10.1521/jscp.2014.33.8.701]

5. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. Soc Issues Policy Rev 2017;11(1):274-302. [doi: 10.1111/sipr.12033]

6. Frost RL, Rickwood DJ. A systematic review of the mental health outcomes associated with Facebook use. Comput Human Behav 2017;76:576-600. [doi: 10.1016/j.chb.2017.08.001]

7. Fogg BJ. Persuasive Computers: Perspectives and Research Directions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1998 Presented at: CHI'98; April 18-23, 1998; Los Angeles, California, USA p. 225-232. [doi: 10.1145/274644.274677]

8. Erfani SS, Abedin B. Impacts of the use of social network sites on users' psychological well-being: a systematic review. J Assoc Inf Sci Technol 2018;69(7):900-912. [doi: 10.1002/asi.24015]

9. Ryff CD, Singer BH. Know thyself and become what you are: a eudaimonic approach to psychological well-being. J Happiness Stud 2008;9(1):13-39. [doi: 10.1007/s10902-006-9019-0]

10. Huang C. Internet use and psychological well-being: a meta-analysis. Cyberpsychol Behav Soc Netw 2010 Jun;13(3):241-249. [doi: 10.1089/cyber.2009.0217] [Medline: 20557242]

11. Boyd DM, Ellison NB. Social network sites: definition, history, and scholarship. J Comput Mediat Commun 2007;13(1):210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]

12. Fogg BJ. Persuasive Technology: Using Computers To Change What We Think And Do. San Francisco: Morgan Kaufmann; 2003.

13. Fogg BJ. Persuasive technologies. Commun ACM 1999;42(5):26-29. [doi: 10.1145/301353.301396]

14. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inform Syst 2009;24(1):485-500. [doi: 10.17705/1cais.02428]

15. Centre for Humane Technology. 2019. URL: https://humanetech.com/ [accessed 2019-10-22]

16. Centre for Humane Technology. The Extractive Attention Economy is Tearing Apart Our Shared Social Fabric. URL: http://humanetech.com/problem/ [accessed 2018-11-28]

17. Wayback Machine - Internet Archive. Calling All Technology Makers. URL: http://web.archive.org/web/20181113105139/http://humanetech.com/designers/ [accessed 2019-09-20]

18. Diener E. Subjective well-being. Psychological Bulletin 1984;95(3):542-575. [doi: 10.1007/978-90-481-2350-6_2]

19. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: three decades of progress. Psychol Bull 1999;125(2):276-302. [doi: 10.1037//0033-2909.125.2.276]

20. Ryan RM, Deci EL. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 2001;52:141-166. [doi: 10.1146/annurev.psych.52.1.141] [Medline: 11148302]

21. Ryff CD, Singer BH. Best news yet on the six-factor model of well-being. Soc Sci Res 2006;35(4):1103-1119. [doi: 10.1016/j.ssresearch.2006.01.002]

22. Ryff CD. Psychological well-being revisited: advances in the science and practice of eudaimonia. Psychother Psychosom 2014;83(1):10-28 [FREE Full text] [doi: 10.1159/000353263] [Medline: 24281296]

23. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]

24. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial intelligence-assisted online social therapy for youth mental health. Front Psychol 2017;8:796 [FREE Full text] [doi: 10.3389/fpsyg.2017.00796] [Medline: 28626431]

25. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]

XSL•FO

RenderX

26. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy: designing and evaluating technology for mental health. ACM Trans Comput-Hum Interact 2014;21(1):1-26. [doi: 10.1145/2513179]

27. Peterson C. A Primer in Positive Psychology. New York: Oxford University Press; 2006.

28. Seligman ME, Csikszentmihalyi M. Positive psychology: An introduction. Am Psychol 2000;55(1):5-14. [doi: 10.1037//0003-066x.55.1.5]

29. Wadley G, Lederman R, Gleeson J, Alvarez-Jimenez M. Participatory Design of an Online Therapy for Youth Mental Health. In: Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration. 2013 Presented at: OZCHI'13; November 25–29, 2013; Adelaide, SA, Australia. [doi: 10.1145/2541016.2541030]

30. Gleeson J, Lederman R, Koval P, Wadley G, Bendall S, Cotton S, et al. Moderated online social therapy: a model for reducing stress in carers of young people diagnosed with mental health disorders. Front Psychol 2017;8:485 [FREE Full text] [doi: 10.3389/fpsyg.2017.00485] [Medline: 28421012]

31. Calvo RA, Peters D. Positive Computing: Technology for Wellbeing and Human Potential. Cambridge: MIT Press; 2014.

32. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]

33. Alvarez-Jimenez M, Gleeson J, Bendall S, Penn D, Yung A, Ryan R, et al. Enhancing social functioning in young people at Ultra High Risk (UHR) for psychosis: a pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]

34. Gleeson J, Lederman R, Herrman H, Koval P, Eleftheriadis D, Bendall S, et al. Moderated online social therapy for carers of young people recovering from first-episode psychosis: study protocol for a randomised controlled trial. Trials 2017 Jan 17;18(1):27 [FREE Full text] [doi: 10.1186/s13063-016-1775-5] [Medline: 28095883]

35. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104 [FREE Full text] [doi: 10.1136/bmjopen-2018-024104] [Medline: 30782893]

36. headspace National Youth Mental Health Foundation. Welcome to eheadspace. URL: https://headspace.org.au/eheadspace/ [accessed 2019-10-22]

37. Rice S, Gleeson J, Leicester S, Bendall S, D'Alfonso S, Gilbertson T, et al. Implementation of the enhanced Moderated Online Social Therapy (MOST+) model within a national youth e-Mental health service (eheadspace): protocol for a single group pilot study for help-seeking young people. JMIR Res Protoc 2018 Feb 22;7(2):e48 [FREE Full text] [doi: 10.2196/resprot.8813] [Medline: 29472177]

38. van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]

39. Peterson C, Seligman M. Character Strengths and Virtues: A Handbook and Classification. Oxford, UK: Oxford University Press; 2004.

40. Seligman ME. Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment. New York: Free Press; 2002.

41. Nezu AM, Nezu CM, D'Zurilla T. Problem-Solving Therapy: A Treatment Manual. New York, NY: Springer Publishing Co; 2013.

42. Kuss DJ, Griffiths MD. Online social networking and addiction--a review of the psychological literature. Int J Environ Res Public Health 2011 Sep;8(9):3528-3552 [FREE Full text] [doi: 10.3390/ijerph8093528] [Medline: 22016701]

43. Kuss D, Griffiths M. Social networking sites and addiction: ten lessons learned. Int J Environ Res Public Health 2017 Mar 17;14(3):pii: E311 [FREE Full text] [doi: 10.3390/ijerph14030311] [Medline: 28304359]

44. Wozney L, Huguet A, Bennett K, Radomski AD, Hartling L, Dyson M, et al. How do eHealth programs for adolescents with depression work? A realist review of persuasive system design components in internet-based psychological therapies. J Med Internet Res 2017 Aug 9;19(8):e266 [FREE Full text] [doi: 10.2196/jmir.7573] [Medline: 28793983]

45. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

46. D'Alfonso S, Carpenter N, Alvarez-Jimenez M. Making the MOST Out of Smartphone Opportunities for Mental Health. In: Proceedings of the 30th Australian Conference on Computer-Human Interaction. 2018 Presented at: OzCHI'18; December 4-7, 2018; Melbourne, Australia p. 577-581. [doi: 10.1145/3292147.3292230]

47. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. Using Game-design Elements in Non-gaming Contexts. In: Proceedings of the Extended Abstracts on Human Factors in Computing Systems. 2011 Presented at: CHI EA'11; May 7-12, 2011; Vancouver, BC, Canada p. 2425-2428. [doi: 10.1145/1979742.1979575]

48. Llagostera E. On Gamification and Persuasion. In: Proceedings of SBGames 2012. 2012 Presented at: SBC'12; November 2-4, 2012; Brasilia, Brazil.

49.   Hamari J, Koivisto J. Social Motivations To Use Gamification: An Empirical Study Of Gamifying Exercise. In: Proceedings of the 21st European Conference on Information Systems. 2013 Presented at: ECIS'13; June 5-8 2013; Utrecht.

50.   Selinger E, Sadowski J, Seager T. Gamification and morality. In: Walz SP, Deterding S, editors. The Gameful World: Approaches, Issues, Applications. Cambridge: MIT Press; 2014:371-392.

51.   Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]

52.   Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]

53.   Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 4;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]

54.   Pereira P, Duarte E, Rebelo F, Noriega P. A review of gamification for health-related contexts. In: Marcus A, editor. Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments. Switzerland: Springer; 2014:742-753.

55.   Kim TW, Werbach K. More than just a game: ethical issues in gamification. Ethics Inf Technol 2016;18(2):157-173. [doi: 10.1007/s10676-016-9401-5]

56.   Cavusoglu H, Zhuolun L, Huang K. Can Gamification Motivate Voluntary Contributions?: The Case of StackOverflow Q&A Community. In: Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing. 2015 Presented at: CSCW'15 Companion; March 14-18, 2015; Vancouver, BC, Canada p. 171-174. [doi: 10.1145/2685553.2698999]

57.   Bornfeld B, Rafaeli S. Gamifying with badges: a big data natural experiment on Stack Exchange. First Monday 2017;22(6). [doi: 10.5210/fm.v22i6.7299]

58.   FeverBee. 2017 Dec 17. The Most Advanced Online Community On The Web (A breakdown Of StackOverflow's Online Community. URL: https://www.feverbee.com/stackoverflow/ [accessed 2018-11-27]

59.   Atwood J. Stack Overflow. 2009 Mar 9. The Value of Downvoting, or, How Hacker News Gets It Wrong. URL: https://stackoverflow.blog/2009/03/09/the-value-of-downvoting-or-how-hacker-news-gets-it-wrong/ [accessed 2018-11-27]

60.   Knutas A. Gamification. In: Lee N, editor. Encyclopedia of Computer Graphics and Games. Cham, Switzerland: Springer; 2018.

Abbreviations

CHT: Center for Humane Technology
HCI: human-computer interaction
MOST: Moderated Online Social Therapy
PWB: psychological well-being
RCT: randomized controlled trial
SDT: Self-Determination Theory
SNS: social networking site
SWB: subjective well-being
TIO: Talk It Out

XSL•FO

RenderX

XSL•FO
RenderX