A review of popular smartphone apps for depression and anxiety: Assessing the inclusion of evidence-based content

Akash R. Wasil*,1, Katherine E. Venturo-Conerly, Rebecca M. Singleton, John R. Weisz
Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA, 02138, USA

ARTICLE INFO

Keywords:
Digital health
mHealth
Telemedicine
Mobile health
Psychological treatment
Evidence-based practice
Smartphone apps
Depression
Anxiety

ABSTRACT

Smartphone applications for the treatment of depression and anxiety have acquired millions of users, yet little is known about whether they include evidence-based therapeutic content. We examined the extent to which popular mental health applications (MH apps) for depression and anxiety contain treatment elements found in empirically supported psychotherapy protocols (i.e., “common elements”). Of the 27 MH apps reviewed, 23 included at least one common element, with a median of three elements. Psychoeducation (in 52% of apps), relaxation (44%), meditation (41%), mindfulness (37%), and assessment (37%) were the most frequent elements, whereas several elements (e.g., problem solving) were not found in any apps. We also identified gaps between app content and empirically supported treatments. Cognitive restructuring was more common in depression protocols than in depression apps (75% of protocols vs. 31% of apps), as was problem solving (34% vs. 0%). For anxiety, exposure (85%, 12%), cognitive restructuring (60%, 12%), and problem solving (25%, 0%) were more common in protocols than apps. Overall, our findings highlight empirically supported treatment elements that are poorly represented in current MH apps. The absence of several core treatment elements underscores the need for future research, including randomized trials testing the effectiveness of popular MH apps.

1. Introduction

If you type in “depression” [in the app store], it’s hard to know if the apps that you get back are high quality, if they work, if they’re even safe to use …. Right now it almost feels like the Wild West of health care.”—John Torous, quoted in Nature (Anthes, 2016, p. 21)

Although evidence-based interventions exist for many mental health problems, most people in need of treatment do not receive it (Kazdin & Blase, 2011). Common barriers include high costs for treatment, a dearth of trained clinicians in many areas, stigma associated with seeking help, transportation and scheduling difficulties, and personal preferences for self-help (Andrade et al., 2014; Gulliver, Griffiths, & Christensen, 2010). Recent advances in technology could allow for a dramatic expansion in access to mental health care (Fairburn & Patel, 2017; Kazdin, 2017). Mental health smartphone applications (MH apps) may offer less stigmatizing treatment options, especially for individuals who prefer self-help. MH apps can also be flexible (Anthes, 2016) and cost-effective (Paganini Teigelkötter, Buntrock, & Baumeister, 2018), potentially improving the reach and dissemination of evidence-based mental health care (Agras, Fitzsimmons-Craft, & Wilfley, 2017; Fairburn & Patel, 2017; Kazdin, 2017). Indeed, internet-based interventions have been effective for a variety of mental health conditions, (Andrews, Cuijpers, Craske, McEvoy, & Titov, 2010; Ebert et al., 2015; Kaplan & Stone, 2013), and recent meta-analyses suggest that interventions delivered via smartphones can help treat depression (Josephine, Josefine, Phillip, David, & Baumeister, 2017) and anxiety (Firth et al., 2017).

However, a significant “digital research-practice gap” may exist in the field of digital mental health. Digital interventions that are empirically tested and supported are not necessarily widely used, and those that are popular may not have been tested in credible ways (Firth et al., 2017a, 2017b; Torous et al., 2018). Moreover, of those that do become available often fail to attract or retain users (Christensen, Griffiths, Groves, & Korten, 2006; Fleming et al., 2016, 2017). For example, the average total time spent on PTSD Coach, an empirically supported MH app for PTSD, was only 5 min (Owen et al., 2015), even though frequent and consistent use is important for treatment gains (Kuhn et al., 2014). In contrast, some interventions found in app stores have attracted and retained millions of users (e.g.,...
useful tools in the dissemination of evidence-based consistent with empirically supported interventions. This issue is quire requirements for MH apps on the Apple App Store or Google Play Store (Donker et al., 2013). Headspace), yet little is known about the effectiveness of most of these commercially successful MH apps (Donker et al., 2013).

Given the lack of scientific evidence, the absence of scientific requirements for MH apps on the Apple App Store or Google Play Store (Anthes, 2016), and limitations in existing app evaluation methods (Torous et al., 2018), relatively little is known about the effectiveness of popular MH apps or even the extent to which MH apps include content consistent with empirically supported interventions. This issue is especially important given the widespread hope that MH apps will be useful tools in the dissemination of evidence-based content. Indeed, while technology-based interventions could reduce the burden of mental illness (Fairburn & Patel, 2017), the inclusion of empirically supported content would enhance the potential benefit (Kazdin, 2017).

In order to assess the impact of current technology, it will be important to examine the extent to which the content of digital mental health interventions corresponds to content that has been shown to improve mental health outcomes. Fortunately, there is extensive scientific literature identifying interventions that have been tested in randomized controlled trials and found to be effective (e.g., Cuipers, Cristea, Karyotaki, Rejinders, & Huibers, 2016; Nathan & Gorman, 2015; Weisz et al., 2017; Weisz & Kazdin, 2017). A step toward developing hypotheses and questions about the mental health benefits of apps might be an examination of the degree of correspondence between the apps and the content within these empirically supported interventions.

One approach to this process could involve applying the “distillation and matching model” (Chorpita, Daleiden, & Weisz, 2005), also known as the “common elements” approach (Chorpita, Becker, & Daleiden, 2007). This approach involves identifying discrete procedures in empirically supported treatment protocols, referred to as “practice elements” (Chorpita et al., 2005, p. 11). Practice elements that are commonly used in empirically supported treatment protocols are referred to as “common elements” (Chorpita et al., 2007, p. 647). For example, protocols for depression could be broken down into common elements such as cognitive restructuring, behavioral activation, and relaxation. Applying the distillation and matching model, researchers have identified numerous common elements in empirically supported treatment protocols for depression and anxiety in young people, as well as common treatment engagement strategies for youth and adults (Chorpita & Daleiden, 2009; Higa-McMillan, Francis, Rith-Najarian, & Chorpita, 2016; Lindsey et al., 2014).

The common elements approach might be applied to MH apps in order to gauge how many elements in an app correspond to those found in empirically supported mental health interventions. Notably, the presence of evidence-based treatment elements would not, in and of itself, guarantee that an app would be effective. However, the elements approach could provide a useful metric as a first step toward judging the potential effectiveness of an intervention that has not been empirically tested. Although several researchers have applied this technique to psychotherapy manuals (Chorpita & Daleiden, 2009; Higa-McMillan et al., 2016; Lindsey et al., 2014), we are not aware of any previous efforts to identify evidence-based elements within MH apps.

Applying the distillation and matching model to popular MH apps may be especially useful, as the content within these apps has been widely disseminated. Furthermore, despite the important contributions from previous reviews of MH apps, these reviews have rarely looked at the apps most commonly encountered in everyday use—i.e., those that have reached hundreds of thousands of users. Instead, the reviews of MH apps have generally focused on a specific subset of apps, potentially excluding popular MH apps that do not fall within the inclusion criteria. For instance, one review included MH apps developed by research institutions and apps that included references to scientific literature (Coulon, Monroe, & West, 2016), another reviewed a handful of commercially available apps yet focused on those that had been tested in scientific trials (Van Ameringen, Turna, Khaleesi, Pullia, & Patterson, 2017), and two recent reviews focused exclusively on mindfulness apps (Mani, Kavanagh, Hides, & Stoyanov, 2015; Plaza, Demarzo, Herrera-Mercadal, & García-Campayo, 2013). The purpose of these reviews was to advance knowledge about specific subsets of MH apps, and such reviews can be valuable in a number of ways; however, their strict inclusion criteria could rule out some of the most popular and widely disseminated MH apps. Given that one of the greatest advantages of MH apps is their potential to disseminate content beyond the scope of face-to-face interventions (Fairburn & Patel, 2017; Kazdin, 2017), it may be particularly useful to focus attention on those MH apps that have achieved widespread use (i.e., popular MH apps), assessing the extent to which the content that has appealed to the largest number of users corresponds to the content of empirically supported interventions.

Additionally, many previous reviews have comprehensively searched app stores to identify all available MH apps that meet a set of inclusion criteria (e.g., Coulon et al., 2016; Huguet et al., 2016). This methodology is sensible, especially when the goal is to draw conclusions about what is available in the commercial marketplace. However, if the goal is to characterize the content that users are most frequently encountering on app stores, a comprehensive review may not be appropriate. Although there may be hundreds of apps for a given condition, consumers are unlikely to comprehensively scan through all available options before selecting apps. In fact, in one study, participants spent just over 2 min searching before making a selection (Dogruel, Joeckel, & Bowman, 2015). Given this observation, reviews that include all available apps may offer less utility than reviews that focus on apps that appear within the top search results (see Bry, Chou, Miguel, & Comer, 2018).

In this paper, we a) assess the extent to which evidence-based treatment elements are present within popular smartphone apps for depression and anxiety, b) examine the frequency of specific evidence-based elements within the apps, and c) compare the frequency of treatment elements in MH apps to the frequency of treatment elements in evidence-based treatment protocols. Our approach expands on past literature in a several ways. First, while past research on treatment elements has only examined a few techniques, we assessed the presence or absence of 26 common evidence-based treatment elements based on a comprehensive review of past research and treatment protocols. Second, we designed our search strategy and inclusion criteria to find popular MH apps that are most likely to be encountered by consumers in everyday use. Third, we focused our review on apps for depression and for anxiety, the two most common categories of mental disorder. Fourth, our review is the first to quantify specific gaps between content in MH apps and content in scientifically supported treatment protocols.

2. Method

2.1. Search strategy & inclusion criteria

We searched the Apple App Store and the Google Play Store in July of 2018. We conducted a search using the term “depression” and a search using the term “anxiety.” Then, we identified additional search terms commonly used by app consumers. While conducting the search, both the Apple App Store and Google Play Store recommend related search terms in a drop-down menu. We used the first three of these recommended search terms to emulate the experience of an ordinary app consumer. On the Apple App Store, the recommended search terms related to depression were “depression games,” “depression helper,” and “depression tracker”; the recommended search terms related to anxiety were “anxiety relief,” “anxiety relief games,” and “anxiety & panic attacks.” In the Google Play Store, the recommended search terms related to depression were “depression wallpapers,” “depression games,” and “depression and anxiety”; the recommended search terms related to anxiety were “anxiety relief apps,” “anxiety helper,” and “anxiety relief games.” This meant that, in total, we used four depression-related search terms and four anxiety-related search terms on each app store.
For each search term listed above, we examined the top five apps from the search results. We limited our search in order to identify popular MH apps that would be most frequently encountered and used by app consumers. Apps that appear more frequently in the search results tend to be more popular and more often used (Apple, n.d.; Peris, n.d.), and app users rarely scroll past the first five apps (Dagruel et al., 2015).

Seeking to assess MH apps that help-seekers naturally encounter and download, we applied relatively few exclusion criteria. We excluded MH apps that were clearly not designed to treat or assess a mental health condition (e.g., Depression Wallpaper, an app that provides screen backgrounds with messages such as “Here’s to all the broken people.”). We also excluded apps that were not available for free, as paid apps are much less likely to be downloaded than free apps (Dagruel et al., 2015). For apps that offered free versions and premium versions, we coded the content that was available in the free version of the app. Similarly, for apps with optional in-app purchases or upgrades, we only coded the free content. The first author applied the inclusion criteria to all apps identified from the search. Then, the second author rescreened apps that were excluded by the first author to confirm the decision to exclude these apps. In the event that an app’s title, description, or content was ambiguous, discussions between the first three authors were used to reach a final judgment.

2.2. Selection of evidence-based treatment elements

We developed a codebook of 26 evidence-based treatment elements. A treatment element has been defined as:

A discrete clinical technique or strategy (e.g., “time out,” “relaxation”) used as part of a larger intervention plan (e.g., a manualized treatment program for youth depression), based on the assumptions that (a) practice elements can be explicitly defined (e.g., using a definition or coding manual), (b) their presence within various interventions can be reliably coded, and (c) different treatments may share practice elements in common. (Chorpita et al., 2005, p. 11)

To develop our list of treatment elements, we reviewed the existing literature on common treatment elements. Several past studies have reviewed empirically supported treatment manuals to identify and define common practice elements in evidence-based interventions for youth and adolescents (Chorpita & Daleiden, 2009; Higa-McMillan et al., 2016; Lindsay et al., 2014). We applied codes from these studies (Chorpita & Daleiden, 2009; Higa-McMillan et al., 2016) that were identified in at least 10% of evidence-based treatment protocols for depression or anxiety. To identify practice elements that were identified in evidence-based protocols for youth depression, we used the data presented by Chorpita and Daleiden (2009). To identify practice elements that were identified in evidence-based protocols for youth anxiety, we used the data presented by Chorpita and Daleiden (2009) and the recent update by Higa-McMillan et al. (2016). In order to maximize the comprehensiveness of our codebook, we required an element to be present in 10% or more of the protocols reviewed in either the Chorpita and Daleiden (2009) or the more recent review by Higa-McMillan et al. (2016). We also used the review by Lindsay and colleagues (2014) to include elements that may promote engagement among digital interventions. Finally, we excluded three treatment elements that were irrelevant to the included digital interventions (see Supplementary File for details).

Because these reviews of treatment elements only examined child and adolescent psychotherapy research, we performed our own review of relevant protocols to identify treatment elements from the adult literature. This was intended to ensure that key elements from adult protocols would be represented in our codebook. First, we reviewed meta-analyses and reviews of evidence-based interventions for adults (Chambless & Hollon, 1998; Chambless & Ollendick, 2001; Cuipers et al., 2008, 2013). From this literature, we identified three treatment modalities with clear empirical support for adult depression and anxiety: cognitive-behavioral therapy (CBT), interpersonal therapy (IPT), and acceptance and commitment therapy (ACT). We then identified treatment elements within each of these treatment modalities. The first and second authors independently reviewed one treatment manual for each modality: for CBT, The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders (Barlow et al., 2011); for IPT, The Guide to Interpersonal Psychotherapy (Weissman, Markowitz, & Klerman, 2017); for ACT, The Mindfulness and Acceptance Workbook for Depression (Strosahl & Robinson, 2017). Next, the first and second authors each independently reviewed the fourth edition of Treatments That Work (Nathan & Gorman, 2015), a book synthesizing research on the latest empirically supported treatments almost exclusively for adults. Treatments That Work lists empirically supported treatments for adult depression (chapter 13) and adult anxiety (chapter 14) and describes elements within these treatments (Nathan & Gorman, 2015). In addition, the first and second authors also independently reviewed articles by the relevant treatment developer teams that were cited in Treatments That Work for each empirically supported treatment modality: behavioral therapy for depression (Dimidjian, Barrera, Martell, Muñoz, & Lewinsohn, 2011), cognitive-behavioral therapy for depression (Dobson et al., 2008), interpersonal therapy for depression (Schramm et al., 2007), cognitive-behavioral therapy for anxiety (Dugas et al., 2010), and social skills training for anxiety (Juster & Heimberg, 1995). While reviewing these sources, the first and second authors independently identified treatment elements that were not already identified by the review articles focused on youth psychotherapy protocols (Chorpita & Daleiden, 2009; Higa-McMillan et al., 2016; Lindsay et al., 2014). The first two authors discussed these non-overlapping treatment elements with the third and fourth author to update the codebook. Disagreements were resolved via discussion.

The final codebook consisted of 26 treatment element codes based on our review of previous studies on youth (Chorpita & Daleiden, 2009; Higa-McMillan et al., 2016; Lindsay et al., 2014), our review of treatment manuals, Treatments That Work, and published papers on empirically supported treatments for adults (Barlow et al., 2011; Dimidjian et al., 2011; Dobson et al., 2008; Dugas et al., 2010; Heimberg & Juster, 1995; Nathan & Gorman, 2015; Schramm et al., 2007; Strosahl & Robinson, 2017; Weissman et al., 2017). A complete list of our codes is presented in Table 1. Additional details about our codes and definitions of codes are available in Table S1 of the Supplementary File.

2.3. Coding procedure

Apps were downloaded and coded in September and October of 2018. Coding consisted of examining the presence or absence of each treatment element. The intensity of a treatment element (i.e., how often it was presented in the app) was not evaluated. The first author coded all of the included apps. The second author coded 33% of these apps (n = 9) to assess inter-rater reliability; Cohen’s kappa ranged from 0.73 to 1.0 (Table 2). In addition to treatment elements, we also coded the app’s number of downloads (using download ranges provided by the Google Play Store2) and average review rating (reporting average ratings from the Apple App Store and the Google Play Store).

---

2 Apps are ordered in searches based on a variety of factors. These include downloads, number of ratings, average ratings, app name, app subtitle, app keywords, and app descriptions. For more, see https://developer.apple.com/app-store/search/ and https://www.appaptain.com/blog/optimize-play-store-app/The Apple App Store and Google Play Store also allow app developers to pay for ad campaigns, allowing an app to appear as the first result in a given search. These apps are labeled with an “Ad” marker. When we conducted our searches, we skipped any “Ad” apps. Therefore, an app with an “Ad” label would only be included if it also appeared in the top five search hits organically (i.e., it would have appeared in the top five even without the ad).

2 Downloads for the Apple App Store are not publicly available.
3. Results

3.1. Search results

Our initial searches identified 80 apps (40 from the depression-related searches and 40 from the anxiety-related searches). Of these, 34 apps were excluded as duplicates, 14 were excluded because they were not intervention apps or assessment apps (e.g., Depression Wallpapers), two were excluded because the app only had a paid version, and three were discontinued over the course of the study. (Fig. 1). If an app was available on both the Google Play Store search and Apple App Store search, we rated the version on the Apple App Store. Our final sample included 27 apps (see Table 2).

3.2. Characteristics of included apps

We examined if each app was primarily an intervention app (designed to offer support or treatment) or a mood tracking/assessment app. We also assessed if the app included specific evidence-based treatment elements. These elements included psychoeducation, assessment, relaxation, meditation, mindfulness, cognitive/coping strategies, activity scheduling, self-monitoring, labelling emotions, expressing kindness, exposure, modeling, goal setting, guided imagery, communication skills, communication analysis, skill building, homework assignment, behavioral contracting, assertiveness training, and identifying values. We calculated Cohen's Kappa for the inter-rater reliability of these assessments. We sorted the apps by the number of evidence-based treatment elements and provided a summary of the apps with the highest number of elements.

Table 1
Frequency of evidence-based treatment elements within publicly available MH apps.

<table>
<thead>
<tr>
<th>Treatment Element</th>
<th>Cohen’s Kappa</th>
<th>Frequency in Apps from Depression Searches (n = 16)</th>
<th>Frequency in Apps from Anxiety Searches (n = 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychoeducation</td>
<td>0.78</td>
<td>8 (50%)</td>
<td>9 (53%)</td>
</tr>
<tr>
<td>Assessment</td>
<td>1.0</td>
<td>8 (50%)</td>
<td>3 (18%)</td>
</tr>
<tr>
<td>Relaxation</td>
<td>0.78</td>
<td>6 (38%)</td>
<td>8 (47%)</td>
</tr>
<tr>
<td>Meditation</td>
<td>1.0</td>
<td>6 (38%)</td>
<td>7 (41%)</td>
</tr>
<tr>
<td>Mindfulness</td>
<td>0.73</td>
<td>6 (38%)</td>
<td>5 (29%)</td>
</tr>
<tr>
<td>Cognitive/Coping</td>
<td>1.0</td>
<td>5 (31%)</td>
<td>2 (12%)</td>
</tr>
<tr>
<td>Activity scheduling/Behavioral activation</td>
<td>1.0</td>
<td>5 (31%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>1.0</td>
<td>4 (25%)</td>
<td>4 (24%)</td>
</tr>
<tr>
<td>Labelling emotions</td>
<td>1.0</td>
<td>4 (25%)</td>
<td>2 (12%)</td>
</tr>
<tr>
<td>Expressing kindness (self)</td>
<td>1.0</td>
<td>3 (19%)</td>
<td>2 (12%)</td>
</tr>
<tr>
<td>Expressing kindness (others)</td>
<td>1.0</td>
<td>2 (13%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Exposure</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>2 (12%)</td>
</tr>
<tr>
<td>Modeling</td>
<td>1.0</td>
<td>2 (13%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Goal setting</td>
<td>1.0</td>
<td>2 (13%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Guided imagery</td>
<td>1.0</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Stimulus control</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Motivational Enhancement</td>
<td>1.0</td>
<td>1 (6%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Identifying values</td>
<td>1.0</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Problem solving</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Communication skills</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Communication Analysis</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Skill Building/Behavioral rehearsal</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Homework assignment</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Family/Significant other engagement</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Behavioral contracting</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Assertiveness training</td>
<td>1.0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Table 2
Publicly available MH apps sorted by number of evidence-based treatment elements.

<table>
<thead>
<tr>
<th>App name</th>
<th>Search terms (depression-related, anxiety-related, or both)</th>
<th>Platform</th>
<th>Rating (Apple; Android)</th>
<th>Downloads (Android)</th>
<th>Number of treatment elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wysa</td>
<td>Depression</td>
<td>Both</td>
<td>4.3; 4.5</td>
<td>100K-500K</td>
<td>11</td>
</tr>
<tr>
<td>Pacifica</td>
<td>Depression</td>
<td>Both</td>
<td>4.7; 4.4</td>
<td>500K-1M</td>
<td>11</td>
</tr>
<tr>
<td>MoodTools</td>
<td>Depression</td>
<td>Both</td>
<td>4.4; 4.3</td>
<td>100K-500K</td>
<td>11</td>
</tr>
<tr>
<td>Youper</td>
<td>Depression</td>
<td>Both</td>
<td>4.9; 4.8</td>
<td>100K-500K</td>
<td>7</td>
</tr>
<tr>
<td>Happify</td>
<td>Depression</td>
<td>Both</td>
<td>4.5; 3.7</td>
<td>100K-500K</td>
<td>7</td>
</tr>
<tr>
<td>7 Cups</td>
<td>Depression</td>
<td>Both</td>
<td>4.1; 4.2</td>
<td>500K-1M</td>
<td>6</td>
</tr>
<tr>
<td>SAM</td>
<td>Anxiety</td>
<td>Both</td>
<td>3.9; 4.0</td>
<td>500K-1M</td>
<td>6</td>
</tr>
<tr>
<td>Simple Habit</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.8; 4.7</td>
<td>500K-1M</td>
<td>6</td>
</tr>
<tr>
<td>Headspace</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.9; 4.5</td>
<td>10M-50M</td>
<td>5</td>
</tr>
<tr>
<td>Moodpath</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.7/5; 4.6/5</td>
<td>100K-500K</td>
<td>4</td>
</tr>
<tr>
<td>Calm</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.8; 4.6</td>
<td>10M-50M</td>
<td>4</td>
</tr>
<tr>
<td>Relax Lite</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.7; 3.8</td>
<td>500K-1M</td>
<td>3</td>
</tr>
<tr>
<td>Jitters CBT</td>
<td>Anxiety</td>
<td>Apple Only</td>
<td>3.3; NA</td>
<td>NA</td>
<td>3</td>
</tr>
<tr>
<td>Rootd</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.1; 4.1</td>
<td>10K-50K</td>
<td>3</td>
</tr>
<tr>
<td>Depression Test (Baris Sarer)</td>
<td>Depression</td>
<td>Both</td>
<td>4.3; 3.8</td>
<td>1K-5K</td>
<td>2</td>
</tr>
<tr>
<td>Meditation Game</td>
<td>Depression</td>
<td>Both</td>
<td>4.8; 4.5</td>
<td>500K-1M</td>
<td>2</td>
</tr>
<tr>
<td>PsychApp Free</td>
<td>Depression</td>
<td>Both</td>
<td>Android Only</td>
<td>NA; 4.0</td>
<td>1K-5K</td>
</tr>
<tr>
<td>Moodtracker Social Diary</td>
<td>Depression</td>
<td>Both</td>
<td>4.5; 4.2</td>
<td>100K-500K</td>
<td>1</td>
</tr>
<tr>
<td>Daylio</td>
<td>Depression</td>
<td>Both</td>
<td>4.8; 4.8</td>
<td>5M-10M</td>
<td>1</td>
</tr>
<tr>
<td>DARE—Break Free From Anxiety</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.8; 4.7</td>
<td>100K-500K</td>
<td>1</td>
</tr>
<tr>
<td>Be Okay</td>
<td>Anxiety</td>
<td>Both</td>
<td>4.7; 5.0</td>
<td>500-1000</td>
<td>1</td>
</tr>
<tr>
<td>Depression Test (by Japps Medical)</td>
<td>Depression</td>
<td>Both</td>
<td>Android Only</td>
<td>NA; 3.5</td>
<td>100K-500K</td>
</tr>
<tr>
<td>Anxiety Test</td>
<td>Anxiety</td>
<td>Both</td>
<td>NA; 3.8</td>
<td>50K-100K</td>
<td>1</td>
</tr>
<tr>
<td>Sunset Micro Journal</td>
<td>Depression</td>
<td>Apple Only</td>
<td>NA; 4.6</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>InnerHour</td>
<td>Anxiety</td>
<td>Both</td>
<td>NA; 4.4</td>
<td>100K-500K</td>
<td>0</td>
</tr>
<tr>
<td>Relieve Depression Hypnosis</td>
<td>Depression</td>
<td>Android Only</td>
<td>NA; 4.0</td>
<td>50K-100K</td>
<td>0</td>
</tr>
<tr>
<td>End Anxiety Hypnosis</td>
<td>Anxiety</td>
<td>Android Only</td>
<td>NA; 4.2</td>
<td>100K-500K</td>
<td>0</td>
</tr>
</tbody>
</table>
The analyses above included the intervention apps (n = 22) as well as the assessment or mood-tracking apps (n = 5). Combined analyses are useful, since both kinds of apps are naturally encountered by users and both kinds can offer support to individuals with mental health problems. However, to account for potential differences between intervention apps and assessment or mood-tracking apps, we also conducted separate analyses.

The intervention apps generally had more treatment elements.

---

4Pacific was acquired by Sanvello Health, Inc and has been renamed. The app is now called “Sanvello”. In this paper, we still refer to it as “Pacifica”, as our team coded the content in the app while it was still called Pacifica.
Comparing the content in MH apps to the content of evidence-based treatment protocols for youth, we identified several specific gaps. For instance, cognitive restructuring (22%), activity scheduling/behavioral activation (22%), communication skills (0%), and behavioral contracting (0%) were less common in MH apps than in evidence-based protocols.

Fig. 2. Shows the frequency of evidence-based treatment elements in the depression apps (n = 16).

Table 3
Frequency of practice elements for depression protocols and apps.

<table>
<thead>
<tr>
<th>Practice Element</th>
<th>Frequency in evidence-based treatment protocols for youth</th>
<th>Frequency in MH apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive restructuring</td>
<td>75%</td>
<td>31%</td>
</tr>
<tr>
<td>Psychoeducation</td>
<td>71%</td>
<td>50%</td>
</tr>
<tr>
<td>Activity scheduling/BA</td>
<td>58%</td>
<td>31%</td>
</tr>
<tr>
<td>Problem solving</td>
<td>54%</td>
<td>0%</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>54%</td>
<td>25%</td>
</tr>
<tr>
<td>Goal setting</td>
<td>46%</td>
<td>12%</td>
</tr>
<tr>
<td>Relaxation</td>
<td>42%</td>
<td>38%</td>
</tr>
<tr>
<td>Behavioral contracting</td>
<td>38%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Note: Percentages taken from Chorpita and Daleiden (2009). Their review examined elements within depression protocols tested in randomized controlled trials.

Fig. 3. Shows the frequency of evidence-based treatment elements in the anxiety apps (n = 17).

Table 4
Frequency of practice elements for anxiety protocols and apps.

<table>
<thead>
<tr>
<th>Practice Element</th>
<th>Frequency in evidence-based treatment protocols for youth</th>
<th>Frequency in anxiety apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure</td>
<td>90%</td>
<td>12%</td>
</tr>
<tr>
<td>Cognitive restructuring</td>
<td>60%</td>
<td>12%</td>
</tr>
<tr>
<td>Relaxation</td>
<td>55%</td>
<td>47%</td>
</tr>
<tr>
<td>Psychoeducation</td>
<td>40%</td>
<td>53%</td>
</tr>
<tr>
<td>Modeling</td>
<td>35%</td>
<td>0%</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>30%</td>
<td>24%</td>
</tr>
<tr>
<td>Problem solving</td>
<td>25%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Note: Percentages taken from Higa-McMillan et al., 2016. Their review examined elements within youth anxiety protocols tested in randomized controlled trials.

Comparing the content in MH apps to the content of evidence-based treatment protocols for youth, we identified several specific gaps. For instance, cognitive restructuring (22%), activity scheduling/behavioral activation (22%), and communication skills (0%) were less common in MH apps than in evidence-based protocols.
Serious games are games designed for purposes other than enjoyment. They are common in empirically supported treatment protocols for depression and anxiety. These data suggest that there may be opportunities for researchers and app developers to invest in novel MH apps teaching exposure, problem solving, activity scheduling/behavioral activation, and cognitive restructuring. Although some publicly available MH apps may already teach these skills, our review highlights that few of the popular MH apps incorporate these practice elements. Thus, future research on MH apps teaching these practice elements could incorporate user-centered design principles (see Torous et al., 2018) in order to develop scalable MH apps that disseminate these practice elements. Future research on exposure-based digital mental health interventions may be especially valuable. Exposure is an essential component of treatment for anxiety disorders, included in 90% of empirically supported youth anxiety protocols (Higa-McMillan et al., 2016). However, it is possible that exposure may not be successful or appropriate in digital mediums: exposure requires participants to actively confront stressful situations, exposure has generally been tested in controlled settings by trained therapists, and in-person exposure therapies often have dropout rates above 20% (e.g., Hofmann, 2004). Thus, an emphasis on patient motivation, user engagement, and user safety may be especially important for future research on exposure-based MH apps.

The MH apps varied in the amount of content they provided and the amount of evidence-based treatment elements they included. Some MH apps focused on one particular treatment element (e.g., Depression Test focused on assessment) while others delivered a wide range of treatment elements, offering greater choice and flexibility (e.g., Pacifica, MoodTools). It is important to note that MH apps with more treatment elements are not necessarily “better” than apps with fewer. Some individuals may benefit more from apps that provide a breadth of options, whereas others may prefer apps that focus on one particular skill. Moreover, some MH apps may offer a large number of treatment elements that are not evidence-based, while others may exclusively offer evidence-based elements. Furthermore, unlike manualized therapies, many apps allowed users to select among an array of options. While this approach promotes flexibility and increases a user’s ability to freely navigate the app, too much choice may be problematic. To address this concern, future research could investigate strategies to recommend treatment elements in MH apps based on user characteristics, resembling the approach of modular therapies (Chorpita and Daleiden, 2009; Watkins et al., 2016). This personalized approach could also be applied to direct users to apps that are most relevant for them. Users may benefit from “recommendation apps,” which direct users to relevant apps among a suite of options (Cheung et al., 2018; Mohr et al., 2019). Personalized recommendation apps could help users navigate a crowded app store, allowing them to find apps that meet their specific mental health needs or preferences.

Some results from our search itself are worthy of discussion. For instance, 34 of the 80 search results were duplicates (i.e., apps already identified by a previous search). This finding suggests that a relatively small number of MH apps dominate app store searches, and future research may evaluate the efficacy of these highly popular apps (e.g., Headspace, Calm, Pacifica). Additionally, 21 of the 22 intervention apps were unguided self-help interventions without any human support. Even among the paid versions of apps, only 3 of the 22 apps offered human support (beyond forums or opportunities to chat informally with peers). A recent meta-analysis found that guided self-help interventions have outperformed unguided interventions, and unguided self-help formats did not outperform usual care (Cuijpers, Noma, Karyotaki, Cipriani, & Furukawa, 2019). This suggests that there may be limits to the potential of unguided self-help apps to improve users’ mental health. On the other hand, it is possible that the findings of the meta-analysis, which identified interventions from randomized controlled trials in peer-reviewed articles, may not generalize to the commercial app marketplace. As noted previously, one limitation of existing researcher-developed apps is that they often lack engagement and have high rates of discontinuation. For this reason, human support may be especially necessary to motivate patients to engage with treatment content. This may not be as necessary for commercial apps, especially the most engaging apps that are able to retain users. Unguided forms of support are also more likely to disseminate widely, reflected in the fact that almost all of the interventions among the top search hits were unguided. Future research aiming to improve the scalability of guided apps or improve the effectiveness of unguided apps may be especially useful.

We also found that many apps are transdiagnostic, as they are designed to treat symptoms of multiple disorders (i.e., depression and anxiety) or non-specific symptoms (i.e., sleep problems, health, mental health, hope). Notably, many apps were “adiagnostic,” in that they attempted to appeal to non-clinical populations and explicitly mentioned that anyone could benefit from the app. Furthermore, at the time of our search, “depression games” and “anxiety relief games” were recommended searches on both the Apple App Store and the Google Play Store. It appears that app users are likely to search for gamified MH apps, though research on the effectiveness of gamified apps is limited. Data suggest that computer-delivered serious games are effective interventions for a variety of mental health problems (Merry et al., 2012; Fleming et al., 2017; Cheng, Davenport, Johnson, Vella, & Hickie, 2019), yet research on smartphone-delivered serious games is limited. In fact, a meta-analysis found nine randomized controlled trials of serious games delivered via computers, yet none testing serious games delivered via smartphones (Lau, Smit, Fleming, & Riper, 2017). Taken together, these findings suggest that additional research on gamified interventions and serious games (see Fleming et al., 2017) may be particularly important.

Our study has several limitations. First, we only coded apps at the top of the search results. While this technique allows us to simulate the experiences of real-world users, who are unlikely to search through many options (Dogruel et al., 2015), our findings may not generalize to the app marketplace as a whole. Second, our study is not sufficient to draw causal claims about the factors influencing the effectiveness of MH apps. For instance, certain treatment elements may only be effective when delivered in specific contexts or delivery methods (e.g., face-to-face therapy). Third, our study was not able to compare the frequency of evidence-based treatment elements in MH apps to their frequency in evidence-based treatment protocols for adults, as research on common elements in adult psychotherapy protocols is limited. However, in our review of the adult psychotherapy literature, we encountered many of the same treatment elements as those identified by youth psychotherapy researchers, suggesting that our comparisons between MH apps and evidence-based protocols may still yield useful information for researchers studying adult psychopathology.

Our study also offers important directions for future research. First, we believe that the distillation and matching model (Chorpita et al., 2005) could inform efforts to evaluate commercially available apps. Several organizations are currently employing methods to evaluate the evidence behind commercially available apps in the absence of rigorous randomized trials. For example, PsyberGuide (Neary & Schueller, 2018) provides a Credibility Score for apps based on an app’s level of research support, description of its intervention, number of user ratings, and level of clinical judgment in the development of the app (PsyberGuide, n.d.). Another example is the App Evaluation Model released by the American Psychiatric Association, designed to help consumers and

---

Serious games are games designed for purposes other than enjoyment. They are often designed to improve educational outcomes or health outcomes (see Fleming et al., 2017; Cheng et al., 2019). An example of a serious game is SPARX, a fantasy role-playing game in which players learn CBT techniques to manage depressive symptoms (see Merry et al., 2012).
 clinicians choose appropriate apps (American Psychiatric Association, n.d.). The App Evaluation model has a section on evidence, which encourages clinicians to download apps and examine their face validity. In our view, these approaches could be usefully supplemented by systematic efforts from researchers to examine the content within mental health apps. Benefits of this approach include: 1) It can be employed even when peer-reviewed research on an app is scarce, 2) It may provide a more rigorous evaluation of an app’s content than relying on app descriptions or user ratings, and 3) It does not require clinicians or users to spend their own time downloading and evaluating apps. For these reasons, we believe the distillation and matching model could useful supplement existing approaches to app evaluation.

Second, our findings suggest that future studies could assess the acceptability and efficacy of MH apps using exposure, problem solving, and other uncommonly used treatment elements. Third, future research could apply our coding system to interventions delivered via other digital formats (e.g., internet-delivered interventions) or apps for other types of problems (e.g., eating disorders, autism spectrum disorder). Finally, future research is necessary to understand other features in popular MH apps that may contribute to their effectiveness or popularity (e.g., design features, search engine optimization techniques, advertising strategies).

5. Conclusion

MH apps may play an important role in expanding access to evidence-based mental health services. We found that existing MH apps appear to use some evidence-based practices, a promising finding given that many of these apps have been downloaded hundreds of thousands of times. However, some evidence-based treatment practices are rarely used in MH apps, providing an important direction for research on new MH apps. Importantly, noting the inclusion of evidence-based practices is best regarded as a first step toward evaluating MH apps. Randomized controlled trials testing whether the MH apps actually convey measurable mental health benefits to users are necessary, and such trials should figure prominently in the research agenda for the years ahead.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

The authors acknowledge the Harvard College Research Program for providing funding for this project. The authors also thank Scott Mulligan, Sarah Gillespie, Suh Jung Park, Haley Gelberg, Lois Gelfand, and Michelle Shih for thoughtful advice and edits on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2019.103498.

References


Merry, S. N., Stasiak, K., Shepherd, M., Frampton, C., Fleming, T., & Lucassen, M. F.
Kazdin, A., & Blase, S. (2011). Rebooting psychotherapy research and practice to reduce
Kaplan, R. M., & Stone, A. A. (2013). Bringing the lab and clinic to the community: Mobile
cognitive and
Behavioral Practice, 25(4), 531–537.
Owen, J. E., Jaworski, B. K., Kuhn, E., Makin-Byrd, K. N., Ramsey, K. M., & Hoffman, J. E.
(2015). mHealth in the wild: using novel data to examine the reach, use, and impact of
PTSD coach. JMIR Mental Health, 2(1), e7.
Paganini, S., Teigelkotte, W., Buntrock, C., & Baumeister, H. (2018). Economic evalua-
tions of internet- and mobile-based interventions for the treatment and prevention of
Retrieved from https://www.apptamin.com/blog/optimize-play-store-app/
features. JMIR mHealth and uHealth, 1(2), e24.
psyberguide/#about-credibility.
Schramm, E., van Calker, D., Dykierpek, R., Lieb, K., Keck, S., Zobel, I., ... Berger, M.
(2007). An intensive treatment program of interpersonal psychotherapy plus phar-
macotherapy for depressed inpatients: Acute and long-term results. American Journal
of Psychiatry, 164(5), 768–777.
Streatfiel, K. D., & Robinson, P. J. (2017). The mindful and acceptance workbook for
depression: Using acceptance and commitment therapy to move through depression and
Torous, J., Firth, J., Huckvale, K., Larsen, M. E., Cosco, T. D., Carney, R., ... Keshavan, M.
(2018). The emerging imperative for a consensus approach toward the rating and
clinical recommendation of mental health apps. The Journal of Nervous and Mental
Disease, 206(8), 662–666.
Van Ameringen, M., Tuma, J., Khalesi, Z., Pullia, K., & Patterson, B. (2017). There is an
app for that! the current state of mobile applications (apps) for DSM-5 ob-
cessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood dis-
orders. Depression and Anxiety, 34(6), 526–539.
virtual environments (IMPROVE-2): Study protocol for a phase III trial of the MOST
randomized component selection method for internet cognitive-behavioural therapy
for depression. BMC Psychiatry, 16(1), 345.
Weisz, J. R., Kuppens, S., Ng, M. Y., Eckstein, D., Ugutuo, A. M., Vaughn-Coxum, R.,
et al. (2017). What five decades of research tells us about the effects of youth psy-
chological therapy: A multilevel meta-analysis and implications for science and