ARTICLE ACCEPTED FOR PUBLICATION AT THE *JOURNAL OF CONSULTING AND CLINICAL PSYCHOLOGY*

Reassessing Evidence-Based Content in Popular Smartphone Apps for Depression and Anxiety: Developing and Applying User-Adjusted Analyses

Akash R. Wasil[a], Sarah Gillespie[b], Raveena Patel[a], Annemarie Petre[a], Katherine E. Venturo-Conerly[c], Rebecca M. Shingleton[c], John R. Weisz[c] & Robert J. DeRubeis[a]

[a]Department of Psychology, University of Pennsylvania

[b]Department of Psychology, University of Minnesota

[c]Department of Psychology, Harvard University

Corresponding Author:
Akash R. Wasil
wasil@sas.upenn.edu

Author Note
Dr. Shingleton has received stock options for consulting work for a smartphone application for migraine headache management. The other authors report no conflicts of interest.

Manuscript accepted for publication. **Please cite as:**
Wasil, A.R., Gillespie, S., Patel, R., Petre, A., Venturo-Conerly, K. E., Shingleton, R. M., Weisz, J.R., & DeRubeis, R.J. (2020). Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. *Journal of Consulting and Clinical Psychology*. Advance of Print.

Abstract

Objective: To assess the dissemination of evidence-based content within smartphone apps for depression and anxiety by developing and applying *user-adjusted analysis*—a method for weighting app content based on each app's number of active users.

Method: We searched the Apple App Store and Google Play Store and identified 27 apps within the top search hits, which real-world users are most likely to encounter. We developed a codebook of evidence-based treatment elements by reviewing past research on empirically supported treatments. We coded the apps to develop an initial tally of the frequency of treatment elements within the MH apps. We then developed and applied user-adjusted analysis to refine the tallies based on each app's number of monthly active users.

Results: The two most popular apps were responsible for 90% of monthly active users, and user-adjusted analysis markedly altered conclusions of prior reports based on tallies alone. For example, mindfulness was present in 37% of apps but reached 96% of monthly active users, cognitive restructuring was present in 22% but reached only 2%, and exposure was present in 7% but reached only 0.0004%.

Conclusions: The potential impact of MH apps on mental health may be best evaluated via assessment that combines tallies of evidence-based content with data on the content users are actually accessing. Given wide variation in the popularity of MH apps, findings weighted by usage data may differ markedly from findings based on raw tallies alone.

Keywords: mHealth; depression; anxiety; evidence-based treatment; digital mental health

Public Health Significance Statement

Traditionally, analyses of mobile health and mental health applications have not accounted for the fact that apps differ markedly in their dissemination. In this study, we developed and applied a new method—*user adjusted analysis*—to better understand the content that users are receiving from mobile apps for depression and anxiety. From a public health perspective, this study a) enhances our understanding of mobile apps for depression and anxiety and b) presents a new way of estimating the impact of mobile apps.

Introduction

Most people in need of mental health services do not receive any. Even if the supply of mental health providers in the United States doubled, there would still not be enough service providers to reach most people in need of care (Kazdin & Blase, 2011). Challenges to disseminating mental health care include high costs, a lack of trained providers, a concentration of providers in affluent and urban areas, stigma, transportation and scheduling difficulties, misfit between evidence-based treatments and the mental health delivery system, and preferences for self-help (Andrade et al., 2014; Gulliver, Griffiths, & Christensen, 2010; Weisz, Ng, & Bearman, 2014). Given these challenges, many have called for novel mental health care delivery formats to supplement traditional brick-and-mortar clinics (Fairburn & Patel, 2017; Kazdin, 2017). One particularly promising solution is the creation of mental health applications (MH apps) that can potentially improve the dissemination of evidence-based mental health care (Agras, Fitzsimmons-Craft, & Wilfley, 2017; Fairburn & Patel, 2017; Kazdin, 2017). Enthusiasm about the scalability of MH apps has been accompanied by growing evidence in support of their efficacy. Systematic reviews and meta-analyses suggest that MH apps can be effective for a variety of conditions (Andrews et al., 2010; Ebert et al., 2015; Kaplan & Stone, 2013; Linardon et al., 2019), including depression and anxiety (Firth et al., 2017a; Firth et al., 2017b; Josephine, Josefine, Phillip, David, & Baumeister, 2017; Linardon et al., 2019). Moreover, publicly available apps often include some evidence-based content that is similar to content found in empirically supported psychotherapies (Wasil et al., 2019).

This growing enthusiasm around MH apps has led to several efforts to understand, evaluate, and review existing MH apps. Broadly, these efforts can be categorized in two ways: the analysis of MH apps in peer-reviewed literature and the analysis of publicly available MH apps. Most reviews have focused on apps that have been empirically tested in scientific studies, which are typically apps developed by academic researchers. Less frequently, some investigators have reviewed the content and features within apps that are publicly available, which are typically developed by for-profit companies. Reviews using the first approach have identified MH apps by conducting systematic searches in scientific databases (Bakker et al., 2016; Donker et al., 2013; Firth et al., 2017a; Firth et al., 2017b; Ng, Firth, Minen, & Torous, 2019). These reviews are useful in characterizing the content and efficacy of MH apps that appear in peer-reviewed scientific articles. For instance, meta-analyses of randomized controlled trials suggest that MH apps can be efficacious for a variety of conditions, including depression (Firth et al. 2017a), anxiety and stress (Firth et al. 2017b), and schizophrenia (Firth & Torous, 2015). However, most empirically-tested MH apps included in these meta-analyses are not publicly available (Fleming et al., 2016). Additionally, the empirically supported interventions that are publicly available often struggle to acquire and retain users (Fleming et al., 2017; Mohr, Riper, & Schueller, 2018; Wasil, Gillespie, et al., 2020). As a result, these reviews do not tell us much about publicly available apps that MH app consumers are likely to encounter. This discrepancy between MH apps identified in scientific publications and those that are popular among MH app consumers has been referred to as the “digital research-practice gap” or “research-to-retail gap” (Wasil et al., 2019; Wasil, Gillespie, et al., 2020). Because of this gap, findings about scientifically tested apps may not generalize to publicly available apps. Given this gap, there is a growing need to understand publicly available apps.

Toward this objective, several scholars have searched commercial app stores (e.g., Apple App Store, Google Play Store) and characterized mobile health apps .For example, review articles have been published on publicly available depression and anxiety apps (Van Amerigen et al., 2017; Shen et al., 2015; Wasil et al., 2019), eating disorder apps (Fairburn & Rothwell, 2015), CBT apps (Huguet et al., 2016), mindfulness apps (Mani et al., 2015), stress management apps (Coulon et al., 2016), and physical activity apps (Yang et al., 2015). Generally, these reviews have been systematic and comprehensive—that is, they have searched through hundreds or thousands of apps on the app store in order to find all eligible apps. This approach helps us understand what is *available* via MH apps, but it may not tell us what users *experience* via MH apps, because availability and actual use differ so widely. Indeed, users rarely scroll past the first several apps (Dogruel, Joeckel, & Bowman, 2015). As a result, some authors have tried to limit their inclusion criteria to the top few apps that appear on commercial app searches (Bry, Chou, Miguel, & Comer, 2018; Wasil et al., 2019). Their rationale was that this approach is more likely to focus on the apps that MH app users are likely to encounter. One of these reviews included 121 apps (Bry, Chou, Miguel, & Comer, 2018) and the other included 27 (Wasil et al., 2019). However, even this approach may be insufficient. Recent evidence suggests that there are extreme outlier apps that attract a highly disproportionate share of users. For instance, one study found that the top 3 depression and anxiety apps on the Google Play Store were responsible for over 90% of total monthly active users for depression and anxiety apps, and most apps had zero monthly active users (Wasil et al., 2020).

In light of such findings, reviews that place equal weight on each included app may be unable to draw conclusions about the content that is most commonly accessed by users. In other words, these review articles tell us what content is *available* but may not to tell us what content has actually been *disseminated*. This suggests that the results from systematic reviews of available content may need to be complemented by information on the content that people are commonly accessing. Applying this complementary approach could deepen our understanding of the content to which users are actually exposed via mobile health apps.

Toward this end, we propose user-adjusted analysis—examining app content with weighting to reflect an app's number of active users. In recent years, several organizations have emerged that provide usage estimates of digital health apps (e.g., Mobile Action (https://www.mobileaction.co/). Here, we propose that such data should become a common and core component of reviews of commercial digital health apps. Specifically, we propose that user data be employed to refine the analyses of systematic reviews by weighting analyses based on an app's number of active users. Incorporating user-adjusted analysis to clarify what content users commonly experience should help to better understand the research-to-retail gap. It is possible that user-adjusted analysis may yield conclusions that differ from what unadjusted analyses in systematic reviews have produced.

This logic suggests that user-adjusted analysis could enhance our understanding of mHealth apps, offer novel information about the content that users receive, and help to address the research-to-retail gap. In this study, we describe and apply one type of user-adjusted analysis to refine our understanding of publicly available MH apps for depression and anxiety. Specifically, we analyze the evidence-based content included in MH apps by weighting apps based on their number of monthly number active users. In addition to enhancing our understanding of the content that MH app users receive, our analyses illustrate how user-adjusted analyses can be conducted. In this way, our paper is intended to facilitate use of user-adjusted analyses by other researchers focused on digital health domains.

Method

Search Strategy & Inclusion Criteria

In July of 2018, we searched the Apple App Store and Google Play Store to identify MH apps (see Wasil et al., 2019 for details). First, we conducted one search using the term "depression" and another using the term "anxiety." Then, to supplement these searches using terms and phrases that are commonly used by app consumers, we also applied search terms that were recommended by the dropdown menu of the Apple App Store and Google Play Store. For each store, when we searched "depression," we selected the first three recommended terms/phrases (Apple App Store: "depression games," "depression helper," and "depression tracker"; Google Play Store: "depression wallpapers," "depression games," and "depression and anxiety"). Likewise, we used the same procedure when we searched "anxiety" (Apple App Store: "anxiety relief," "anxiety relief games," and "anxiety & panic attacks"; Google Play Store: "anxiety relief apps," "anxiety helper," and "anxiety relief games"). Thus, we conducted four depression-related searches and four anxiety-related searches on each store.

Because users rarely scroll past the top five apps (Dogruel et al., 2015), we screened the top five apps for each of the 16 searches. We included apps that offer support, treatment content, or assessment techniques (see Wasil et al., 2019 for details). After screening for duplicates (n = 34), we excluded apps if they did not offer treatment content or support (e.g., "Depression Wallpapers"; n = 14), were not free (n = 2), or were discontinued between the initial search and the coding (n = 3). We excluded the two paid apps in order to focus on apps whose access and use would not be altered by potential users' financial resources or willingness to pay. Additionally, previous research on app selection heuristics suggests that people are unlikely to consider downloading paid apps (Dogruel et al., 2015).

After applying these criteria, our final sample included 27 apps. Of these, 10 were identified from the depression-related searches, 11 from the anxiety-related searches, and 6 from both searches.

Coding Evidence-Based Treatment Elements

A detailed description of the process we used to develop and apply our treatment element codebook is available elsewhere (Wasil et al., 2019). In brief, we identified elements that are common among evidence-based psychotherapies for depression and anxiety. We were guided by the distillation and matching model, which describes how manualized treatments can be distilled into discrete practices or treatment elements (see Chorpita, Daleiden, & Weisz, 2005). To identify evidence-based treatment elements, we reviewed existing literature encompassing empirically supported psychotherapies for children and adolescents and for adults. First, we surveyed studies that had used the distillation and matching model to identify treatment elements in youth psychotherapy manuals for depression (Chorpita & Daleiden, 2009) and anxiety (Higa-McMillan et al., 2016). We added elements to our codebook that were present in at least 10% of the evidence-based protocols for youth psychotherapy that these authors surveyed. Next, to supplement elements identified in youth psychotherapy, we completed our own search for elements that are common among adult treatments for depression and anxiety. We surveyed reports by Division 12 of the American Psychological Association (Chambless & Hollon, 1998; Chambless & Ollendick, 2001), which summarized efforts by the American Psychological Association to define and identify empirically supported interventions. To ensure that we included more recent information about empirically supported treatments, we also reviewed meta-analyses (Cuijpers et al., 2008; 2013) and relevant book chapters of *A Guide to Treatments That Work* (Nathan & Gorman, 2015). *A Guide to Treatments that Work* is considered an authoritative source for information about empirically supported treatments for adults; it includes a summary of research on empirically supported treatments and describes the procedures used in these treatments.

Through this process, we identified three treatments with clear empirical support for adult depression and anxiety—cognitive-behavioral therapy (CBT), interpersonal therapy (IPT), and acceptance and commitment therapy (ACT). Then, the first and fifth author reviewed treatment manuals from each modality: CBT (*The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders*; Barlow et al., 2011), IPT (*The Guide to Interpersonal Psychotherapy*; Weissman et al., 2017), and ACT (*The Mindfulness and Acceptance Workbook for Depression*; Strosahl & Robinson, 2017) and independently identified treatment elements that were not already identified in the youth psychotherapy review. Then, these authors discussed a final draft of the codebook with the sixth and seventh authors, two clinical psychologists with experience developing and evaluating treatments for emotional disorders. This process led to a final codebook with treatment elements from both youth and adult psychotherapies for depression and anxiety.

The final version of the codebook included 26 treatment elements (see Supplementary File 1 for details, including a list of treatment elements and their definitions). We coded the presence or absence of each treatment element (see Wasil et al., 2019 for details). Then, we used a random number generator to select a subset (one-third) of the apps. Two authors (AW and KVC) independently coded this subsample of apps (Cohen's kappa ranged from 0.73 to 1.0), and one of these authors (AW) coded the remaining apps.

Usage Data

We collected usage data, including downloads, daily active users (i.e., the average number of unique users who opened the app on a given day), and monthly active users (i.e., the average number of unique users who opened the app over the past month). Daily active users are operationally defined as people who opened an app at least once in a given day, and monthly active users are people who opened an app at least once in a given month.

Data were retrieved from Mobile Action (https://www.mobileaction.co/), a mobile application intelligence platform that tracks real-world app usage. Mobile Action integrates public data (e.g., an app's ranking, star ratings, user reviews) and private data (e.g., information gathered from app developers) to generate these estimates. Data from Mobile Action has been used in previous analyses and offers trends that are consistent with other mobile application intelligence platforms (see Wasil et al., 2020).

Unadjusted vs. User-Adjusted Analysis

To perform our unadjusted analyses, we calculated the frequency of each element in the sample of apps. For instance, imagine that a treatment element was present in nine apps out of a total of twenty-seven apps. The treatment element would be present in nine out of twenty-seven apps (i.e., present in 33% of apps).

To perform our user-adjusted analysis, we weighed each app based on its number of monthly active users. First, we summed the number of monthly active users for the apps containing a certain treatment element. Then, we divided this number by the total number of monthly active users for the entire sample of apps. For instance, imagine that a treatment element was present in nine apps, and these nine apps had a cumulative 100,000 monthly active users out of a total of 1,000,000 monthly active users for the entire sample of apps. The treatment element would be present in apps accounting for 100,000 active users out of 1,000,000 monthly active users (i.e., present in apps responsible for 10% of monthly active users). This procedure allowed us to obtain an estimate of the percentage of monthly active users who are exposed to apps that include each treatment element.

Results

Characteristics of Included Apps

We examined the primary purpose of each app by examining the content in the free version of each app (see Wasil et al., 2019 for details). Twenty-two apps were intervention apps and five were primarily assessment or mood tracking apps. 26 of the 27 included apps did not offer human support; one app (7 Cups) offered peer support. Four apps (Youper, Wysa, 7 Cups, and InnerHour) allowed users to communicate with an AI chatbot.

Usage Data

We examined the number of daily active users and monthly active users for the 27 apps (see Supplementary File 2). Consistent with previous reviews (Baumel, Muench, Edan, & Kane, 2019; Wasil et al., 2020), we collected usage data for a 30-day period. Since our search was conducted in July 2018, we chose the period from July 1, 2018 to August 1, 2018. The apps varied widely in their number of daily active users (*Mean* = 71,130, *SD* = 244,677, *Median* = 315) and monthly active users (*Mean* = 511,721, *SD* = 1,636,153, *Median* = 9,035).

Figure 1 shows the daily active users of each app, and Figure 2 shows the monthly active users. The top 2 apps (Headspace and Calm) accounted for 96% of daily active users and 90% of monthly active users.

Figure 1

Daily Active Users of Mental Health Apps

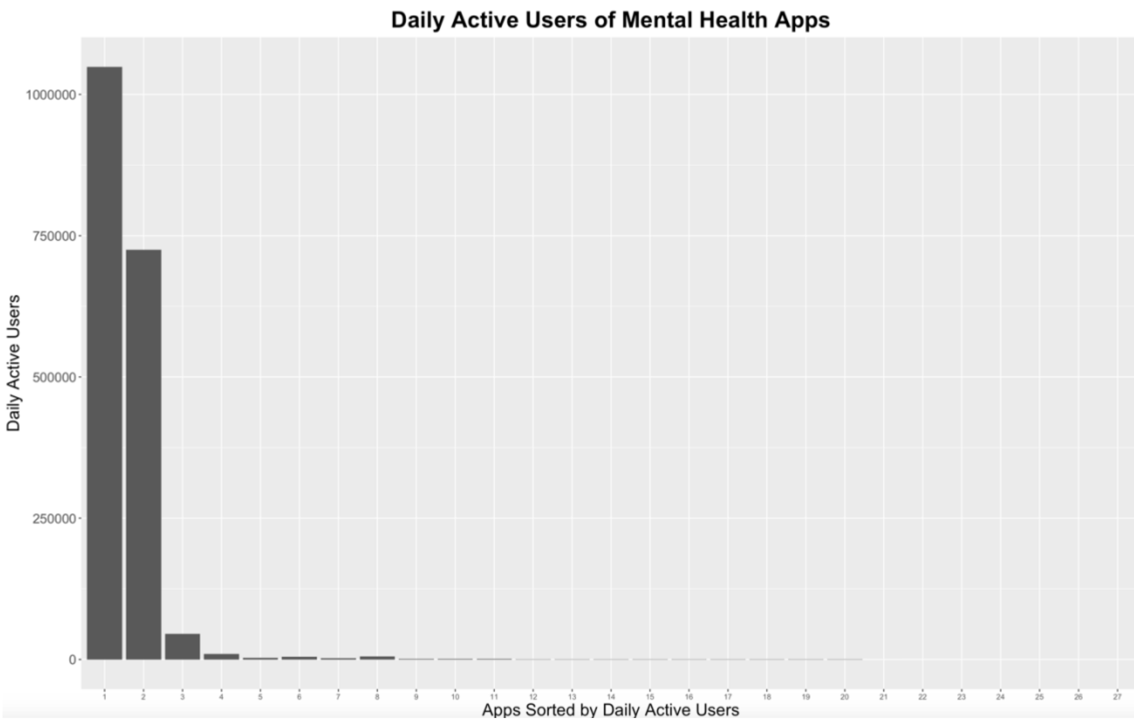

Figure 1. Daily active users of depression and anxiety apps. The top two apps (Headspace and Calm) accounted for 96% of the total number of daily active users. Apps are listed in the following order:

1=Headspace, 2=Calm, 3=Simple Habit- Meditation, 4=Daylio, 5=Wysa, 6=7 Cups, 7=Moodpath: Depression & Anxiety Test, 8=Pacifica, 9=Youper, 10=Happify, 11=InnerHour, 12=DARE- Break Free from Anxiety, 13=Depression Test (Baris Sarer), 14=Sunset Micro Journal, 15=Moodtracker Social Diary, 16=Self-help for Anxiety Management, 17=Relax Lite: Stress and Anxiety Relief, 18=MoodTools- Depression Aid, 19=End Anxiety Hypnosis, 20= Be Okay, 21=Rootd, 22= Meditation Game, 23= Depression Test (Japps Medical), 24=Anxiety Test, 25=Jitters CBT, 26=Relieve Depression Hypnosis, 27=PsychApp Free

Figure 2

Monthly Active Users of Mental Health Apps

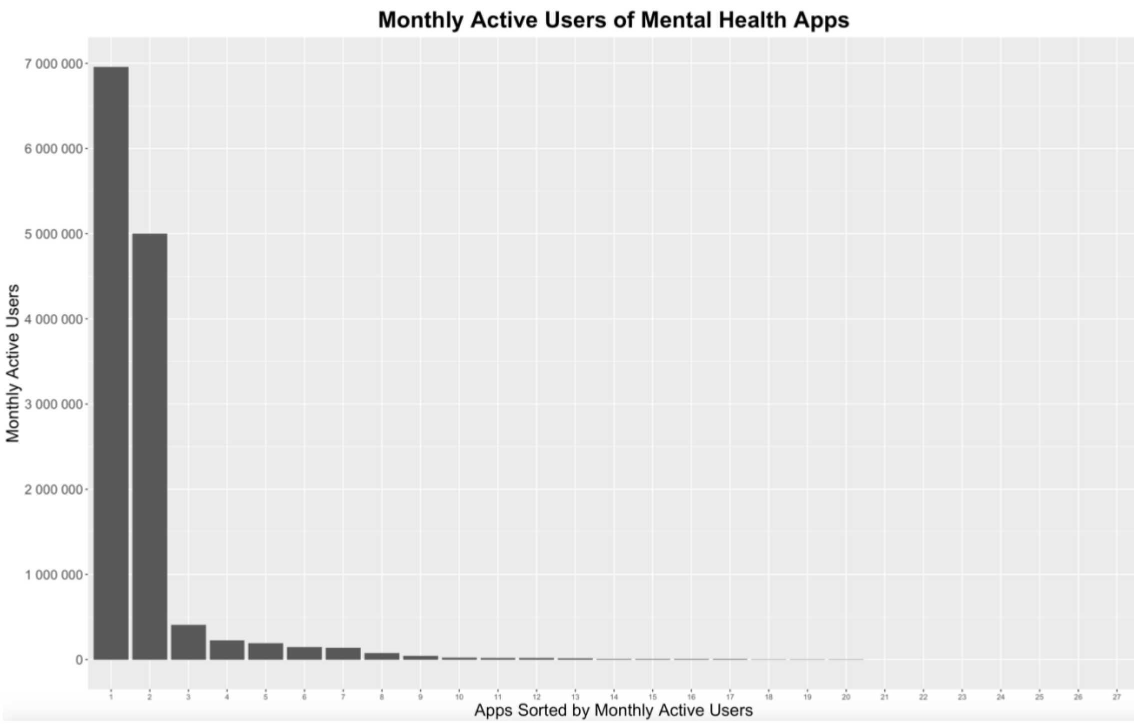

Figure 2. Monthly active users of depression and anxiety apps. The top two apps (Headspace and Calm) accounted for 90% of the total number of monthly active users. Apps are listed in the following order: 1=Headspace, 2=Calm, 3=Simple Habit- Meditation, 4=Daylio, 5=Moodpath: Depression & Anxiety Test 6=7 Cups, 7=Pacifica, 8=Wysa, 9=InnerHour, 10=Happify, 11=Youper, 12=DARE- Break Free from Anxiety, 13=Depression Test (Baris Sarer), 14=Self-help for Anxiety Management, 15=Moodtracker Social Diary, 16=Sunset Micro Journal, 17=MoodTools- Depression Aid, 18=End Anxiety Hypnosis, 19=Relax Lite: Stress and Anxiety Relief, 20=Be Okay, 21=Rootd, 22=Meditation Game, 23=Depression Test (Japps Medical), 24=Anxiety Test, 25=Relieve Depression Hypnosis, 26=Jitters CBT, 27=PsychApp Free

Description of Popular Apps

Given that Headspace and Calm accounted for such a large portion of active users, additional information about these apps could be useful. Both Headspace and Calm are apps that focus on mindfulness and meditation. In both apps, users freely select from a variety of mindfulness and meditation exercises, psychoeducational readings, and animated videos. Users can also set goals like “personal growth,” “work & productivity,” or “stress & anxiety,” and “self-care.” Both apps include user profiles, where users can see information about their app usage (e.g., number of meditations completed, average meditation time, which days they meditated). Among the treatment elements in our codebook, Headspace included five (mindfulness, assessment, crisis management, family/significant other engagement, and stimulus control) and Calm included four (mindfulness, meditation, relaxation, and psychoeducation). Examples of treatment elements within Headspace, Calm, and other MH apps can be found in Supplementary File 3.

Treatment Elements: Unadjusted Analyses and User-Adjusted Analysis

We examined the treatment elements in the apps in two ways. First, we report the frequency of each treatment element in the sample of apps using “unadjusted” analyses that did not take into account an app’s number of active users. Then, we assessed the frequency of each treatment element and adjusted our results based on each app’s number of monthly active users (see Figure 3 and Figure 4).

Figure 3

Unadjusted and User-Adjusted Analysis of Treatment Elements

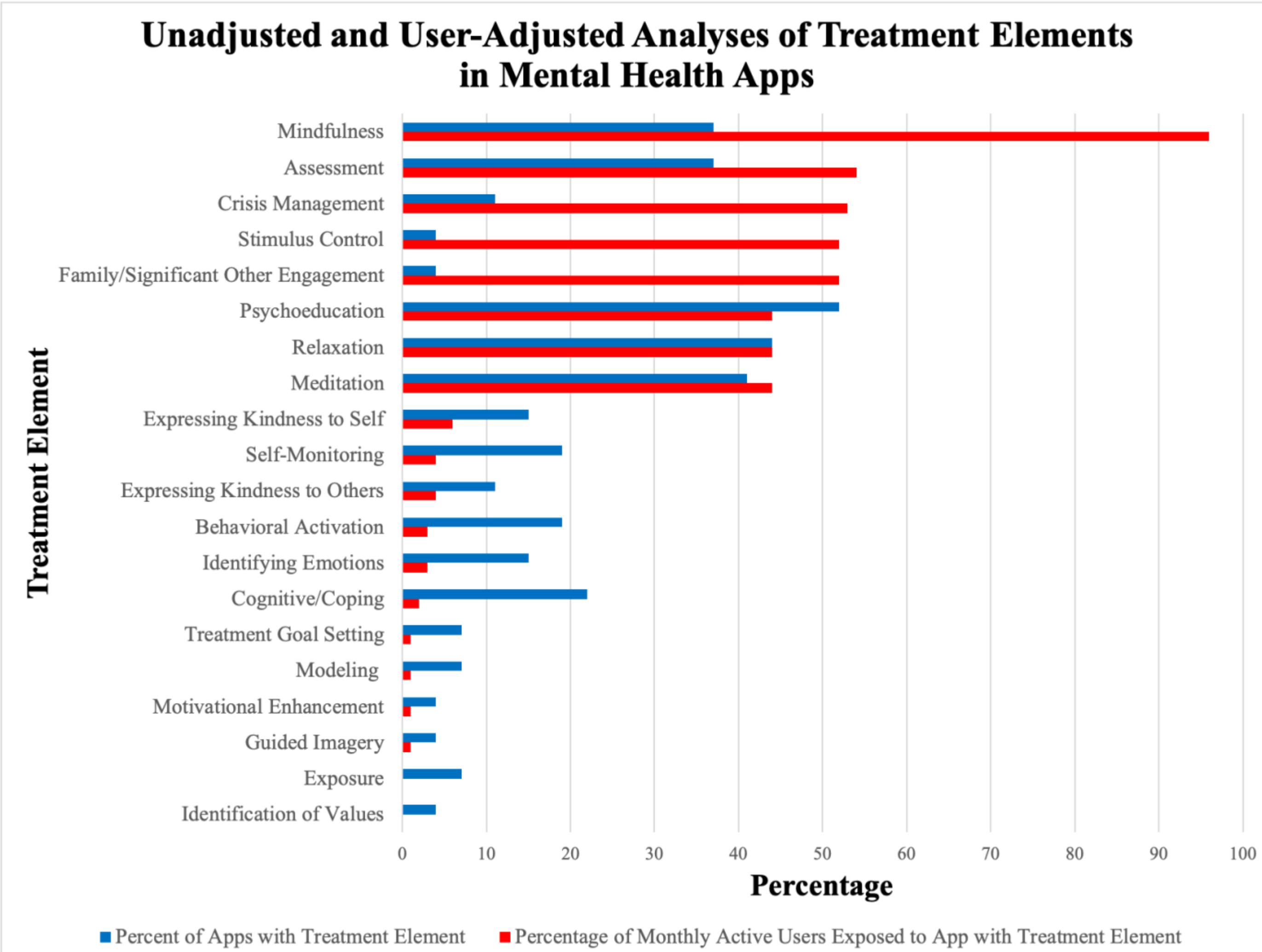

Figure 3. Analysis of treatment elements in depression and anxiety apps. The y-axis shows treatment elements; the x-axis depicts the percentage. Unadjusted analyses, presented in blue, show the percentage of apps that included a given treatment element. User-adjusted analysis, presented in red, show the percentage of monthly active users who used an app that included a given treatment element.

Figure 4

Plot of Unadjusted and User-Adjusted Analyses by Treatment Element

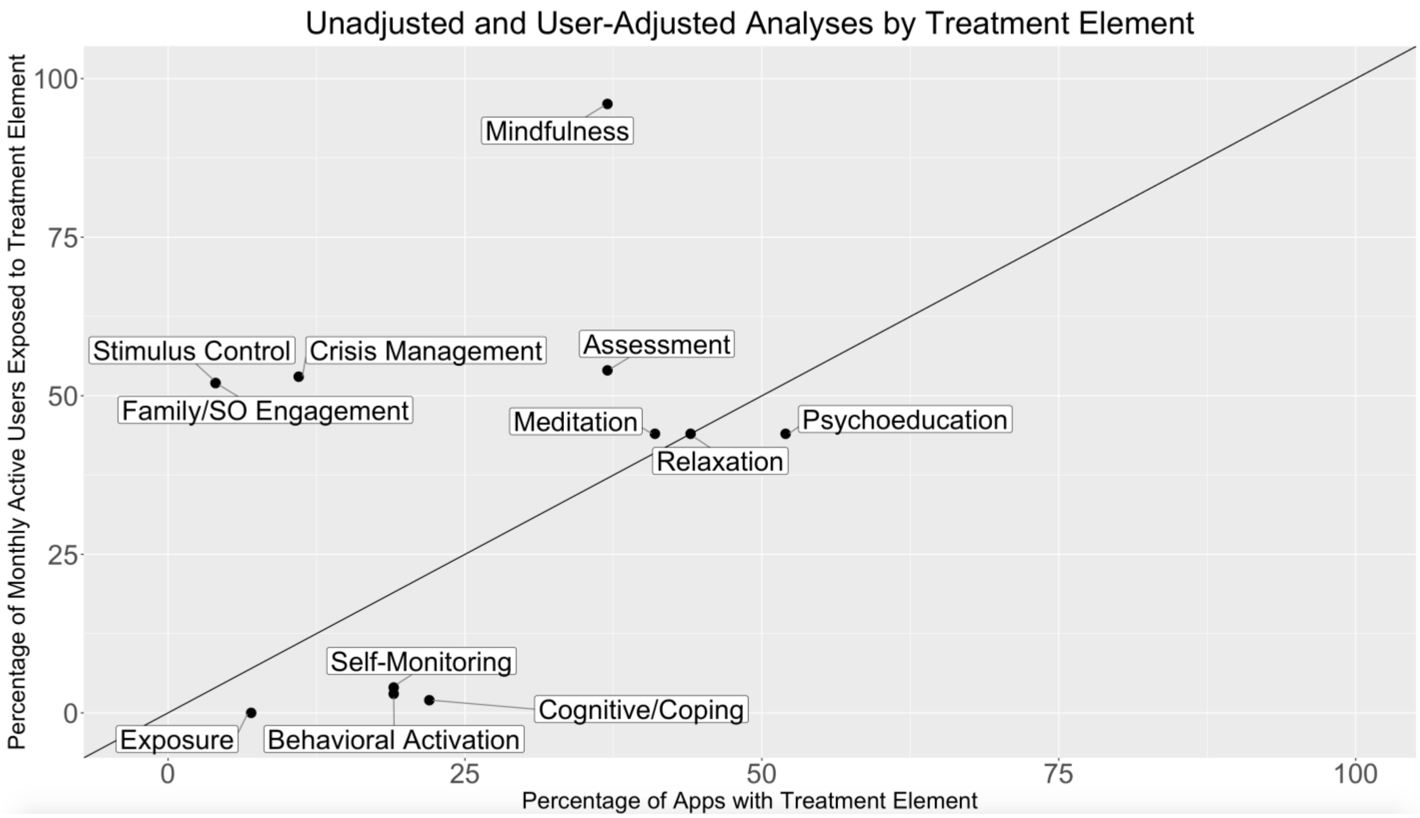

Figure 4. Plot of selected treatment elements in depression and anxiety apps. The points show treatment elements. The x-axis shows the percentage of apps including each treatment element, and the y-axis shows the percentage of monthly active users who used an app with the treatment element. The line has a slope of one, marking where treatment elements would be plotted if the unadjusted analyses (i.e., the percent of apps including the element) were identical to the user-adjusted analysis (i.e., the percent of monthly active users using apps with that element). Treatment elements below the line (e.g., exposure, behavioral activation, cognitive restructuring) are reaching fewer users than unadjusted analyses would suggest, while treatment elements above the line (e.g., mindfulness, crisis management, stimulus control) are likely reaching more users than unadjusted analyses would suggest.

Table 1 lists each treatment element and notes the percentage of apps in which each element appeared (unadjusted analyses) as well as the percentage of monthly active users who accessed an app in which the treatment element appeared (user-adjusted analysis). Because two apps (Headspace and Calm) were responsible for such a large percentage of active users, we performed a sensitivity analysis excluding these apps. When excluding these apps, notable differences between unadjusted and user-adjusted analyses remained, but they were less extreme (see Supplementary File 4).

Table 1

Evidence-Based Treatment Elements Within Publicly Available MH Apps: Unadjusted Analyses and User-Adjusted Analysis

Treatment Element	Frequency in Apps (unadjusted)	Frequency in Apps (weighted by monthly active users)
Psychoeducation	52%	44%
Relaxation	44%	44%
Meditation	41%	44%
Mindfulness	37%	96%
Assessment	37%	54%
Cognitive/Coping	22%	2%
Self-Monitoring	19%	4%

Activity Scheduling/Behavioral Activation	19%	3%
Expressing Kindness to Self	15%	6%
Labelling and Identifying Emotions	15%	3%
Crisis Management	11%	53%
Expressing Kindness to Others	11%	4%
Modeling	7%	1%
Treatment Goal Setting	7%	1%
Exposure	7%	0%
Family/Significant Other Engagement	4%	52%
Stimulus Control	4%	52%
Guided Imagery	4%	1%
Motivational Enhancement	4%	1%
Interoceptive	4%	0%
Identification of Values	4%	0%
Problem Solving	0%	0%

Skill Building/Behavioral Rehearsal	0%	0%
Homework Assignments	0%	0%
Behavioral Contracting	0%	0%
Assertiveness Training	0%	0%

Table 1. A comparison between unadjusted and user-adjusted analysis. Column 1 depicts treatment elements, column 2 depicts the percentage MH apps that included each element, and column 3 depicts the percentage of total monthly active users who used an app that contained each element.

Discussion

We applied user-adjusted analysis to examine the treatment elements in publicly available MH apps. We found that a few highly popular apps (i.e., Headspace and Calm) are responsible for about 90% of monthly active users. According to our estimates, Headspace and Calm had about 7 million and 5 million monthly active users (respectively), several other apps had over 100,000 monthly active users (e.g., Daylio, Moodpath, 7 Cups, Pacifica), most had fewer than 10,000, and some had less than 100. These findings should be interpreted in light of the fact our search procedure and exclusion criteria were designed specifically to identify popular MH apps. For instance, we used search terms recommended by the search engines and only included apps within the top five search hits for each of these terms. Studies that screen a greater number of apps on commercial app stores have found even greater discrepancies between the top apps and the majority of apps (e.g., Wasil et al., 2020). Additionally, we found important differences between unadjusted and user-adjusted analysis. For instance, mindfulness was present in 37% of apps but was included in apps reaching 96% of total monthly active users, family/significant other engagement was included in 4% of apps reaching 52% of monthly active users, cognitive restructuring was present in 22% of apps reaching only 2% of monthly active users, and exposure was present in 7% of apps reaching only 0.0004% of monthly active users. These results illustrate how user-adjusted analyses may enhance reviews of digital health and digital mental health interventions.

Our analyses suggest that some evidence-based treatment elements have disseminated much more widely than others. For instance, mindfulness is included in apps reaching 96% of MH app users; similarly, family/significant other engagement, assessment, meditation relaxation, and crisis management were present in apps reaching about half of MH app users. However, some evidence-based treatment elements are not reaching users. Cognitive restructuring, behavioral activation, exposure, and problem solving—core treatment elements in evidence-based protocols for depression and anxiety (Chorpita & Daleiden, 2009; Higa-McMillan et al., 2016). For instance, cognitive restructuring is present in 75% of empirically supported treatment protocols for youth depression, behavioral activation is present in 58%, and problem solving is present in 54% (Chorpita & Daleiden, 2009). Similarly, exposure is present in 90% of empirically supported treatment protocols for youth anxiety, and cognitive restructuring is present in 60% (Higa-McMillan et al., 2016). Furthermore, a large body of theoretical literature suggests that these elements are active ingredients of empirically supported interventions (Craske, 2010; Foa, 2011; Kendall & Hollon, 2013).

The fact that MH apps are not disseminating several core elements from empirically supported treatments points to an important limitation in the current app marketspace. These findings suggest that researchers and app developers may wish to prioritize the development and dissemination of new apps that include these treatment elements—those that currently are not reaching many users. At the same time, potential app developers should consider our finding that a few MH apps are responsible for over 90% of active users. This finding suggests that the app marketplace functions as a “winner-take-all market” (see Frank & Cook, 2010) in which a handful of highly popular outlier apps attracts nearly all of the attention. These findings suggest that potential app developers should think critically about their ability to engage users and compete in this crowded marketspace before investing resources into new apps (see Wasil, Weisz, & DeRubeis, 2020).

Our findings can also complement previous reviews of MH apps. For example, a recent review of anxiety apps for youth found that 22% of included apps included exposure, and 21% included cognitive restructuring (Bry et al., 2018). Our findings suggest that these elements are reaching fewer users than would be expected based on unadjusted frequency counts. It is possible that *effortful* treatment elements—those that require a great deal of work or effort from MH app users—are less likely to retain a large number of active users. In contrast, treatment elements that provide *immediate relief* (e.g., relaxation, mindfulness) may be more sought out. Exposure may be an especially useful example. Although exposure is a highly evidence-based treatment element for anxiety, exposure can be highly burdensome on participants, and even face-to-face exposure therapies have drop-out rates higher than 20% (Hofmann, 2004). Thus, even though exposure was included in 22% of apps in a previous review (Bry et al., 2018), we predict that the share of users receiving exposure is likely considerably less than 22%. Relatedly, a review of stress management apps concluded that 33% of apps did not include any evidence-based content (Coulon, Monroe, & West, 2016). Our results suggest that this figure may overestimate the number of users using apps without evidence-based content. Given that Headspace and Calm account for over 90% of active users, and both of these apps include evidence-based content, our findings suggest that the number of users receiving evidence-based content may be higher than previously thought. Overall, these examples highlight how user-adjusted analyses may usefully complement the kinds of tallies that are frequently reported in reviews of commercial apps.

Our approach to user-adjusted analyses has a number of important implications for future research. For example, such analyses are useful in understanding what content has disseminated most widely. As discussed, reviews that fail to take into account user data may not be sufficient to understand the content that most people are accessing. Importantly, this implication extends beyond reviews of commercially available apps but also to reviews of apps in peer-reviewed papers. Indeed, numerous reviews and meta-analyses have evaluated the efficacy and features in apps from the peer-reviewed literature (e.g., Firth et al., 2017a; Firth et al., 2017b; Josephine et al., 2017; Weisel et al., 2019). One of these reviews concluded:

> Although some trials showed potential of apps targeting mental health symptoms, using smartphone apps as standalone psychological interventions cannot be recommended based on the current level of evidence. (Weisel et al., 2019, p. 1)

These findings are highly valuable in drawing conclusions about the MH apps that have been empirically tested, though it is worth noting that most of these apps are not available on commercial app stores, and most commercially available apps have not been subject to rigorous empirical testing (Fleming et al., 2016), As a result, it is important not to assume that the findings from MH apps identified in peer-reviewed journal articles will generalize to MH apps “in the wild.” Rather, alternative approaches—including coding the content within these apps, subjecting these apps to randomized controlled trials, and obtaining expert ratings of these apps (e.g., Neary & Schueller, 2018) will be necessary. Given that the majority of users are using a handful of highly popular "outlier apps", future research that aims to understand the potential benefits and harms of these apps will be especially important.

In this way, user-adjusted analyses could be useful in helping to inform app monitoring and evaluation. Several professional societies (e.g., American Psychiatric Association, American Medical Association), government agencies (e.g., US Food and Drug Administration, UK National Health Service), and non-profit organizations (e.g., PsyberGuide; Neary & Schueller, 2018) have released plans to evaluate digital mental health apps. Given that there are thousands of publicly available apps and dozens of important app evaluation frameworks (Henson, David, Albright, & Torous, 2019), it would be costly and time-consuming to comprehensively review all available apps. As a result, efforts to evaluate apps may benefit by prioritizing reviews of highly popular apps that have attracted a large number of active users (e.g., Headspace, Calm). Such reviews could meaningfully inform consumers about the MH apps that they are most likely to use.

Finally, usage data could inform efforts to develop and validate rating scales of app engagement. Popular app rating scales have often been validated by correlating scale ratings with app star ratings (e.g., Stoyanov et al., 2015). A complementary approach would involve using scale ratings to predict active users. While an app's number of users is likely affected by a variety of factors (e.g., marketing and advertising, ease of use, engaging design features), tools that are able to successfully predict app usage or retention data would be highly valuable for investigators. Such an approach would be especially important in light of the poor engagement and dissemination of most digital mental health interventions (Fleming et al., 2016; Fleming et al., 2017; Wasil, Gillespie, et al., 2020).

While a focus on monthly active users is an important strength of our study, our approach also has some limitations. For instance, an app's popularity does not necessarily tell us much about the efficacy of an MH app or its use of evidence-based content. The impact of an MH app is a function of its popularity as well its efficacy. Thus, efforts to evaluate the efficacy of MH apps and the presence of evidence-based content within MH apps are essential. Additionally, our monthly active user figures represent data from just one month (i.e., July of 2018). Future research is needed to understand if usage patterns of MH apps differ month-by-month or season-by-season. Nonetheless, it appears that our findings relating to the distribution of monthly active users (i.e., the fact that a small handful of apps was responsible for the majority of monthly active users) are unlikely to be specific to a given month. We previously found a similar pattern of findings when examining depression and anxiety apps in the month of February of 2019. Headspace and Calm were the most popular MH apps, and they accounted for the majority of active users (Wasil, Gillespie, et al., 2020). Furthermore, using data from January and February of 2020, we found that the top four apps for eating disorders were responsible for 95% of monthly active users (Wasil et al., under review). Therefore, it appears that our findings relating to the distribution of monthly active users is unlikely to be specific to the month of June, and these findings may not even be specific to MH apps for depression and anxiety.

Another potential limitation is that active user estimates may underestimate the impact of apps that are intended for short-term use. Some MH apps may teach skills that users are expected to use and apply without the app. For these apps, once users learn a skill, they may reasonably discontinue the app, thus decreasing the app's number of active users. Finally, apps with high numbers of monthly active users may be attracting a large number of users with subclinical symptoms. Neither Headspace nor Calm is advertised as an app exclusively for people with mental illnesses or mental health problems. On the app store, Headspace claims to offer “mindfulness and stress relief” and Calm claims to offer “mindfulness and sleep stories.” Furthermore, their app store descriptions mention several constructs focused on general health and wellness (e.g., happiness, sleep quality, productivity, exercise, physical health, relationships). This kind of marketing may appeal not only to individuals experiencing diagnosable mental disorders but also to those with subclinical concerns. Our data do not allow us to analyze the proportion of users of MH apps who experience clinical levels of symptoms, and future research is needed to better understand the characteristics of MH app users. Nonetheless, even MH apps that primarily appeal to users with subclinical symptoms could have an important public health impact, as subclinical symptoms of internalizing disorders are associated with a considerable degree of impairment and loss of functioning (Ruscio, 2019).

Importantly, our approach to user-adjusted analyses—weighting app content based on an app’s number of monthly active users—represents only one approach to user-adjusted analyses. Future research could expand on our approach in numerous ways. For instance, such research could examine the number of sessions users spend on MH apps, and the average amount of time they spend per session, an app’s ratio of downloads to active users (i.e., user retention). Currently, investigators are limited by the data that is offered by app analytics organizations. However, over time, it is likely that reliable estimates of these and other metrics will become available, providing more precise information that can inform our understanding of MH Apps. We believe the approach we use in this paper (i.e., weighting based on monthly active users) is one of many valuable ways to incorporate usage data. Ideally, our approach will motivate other investigators to think creatively about how usage metrics can enhance research in a variety of domains.

Overall, we hope that this paper inspires future research incorporating usage data to enhance our understanding of digital mental health interventions. The growing enthusiasm for digital mental health interventions has led to the proliferation of several influential, highly cited review articles. Simple, low-cost strategies that improve the quality of reviews could enhance this growing body of literature examining MH apps. We believe that the inclusion of usage data is one such strategy, offering the potential to meaningfully advance progress in digital health and mental health.

References

Agras, W. S., Fitzsimmons-Craft, E. E., & Wilfley, D. E. (2017). Evolution of cognitive-behavioral therapy for eating disorders. *Behaviour Research and Therapy*, *88,* 26-36.

Andrade, L. H., Alonso, J., Mneimneh, Z., Wells, J. E., Al-Hamzawi, A., Borges, G.. . . Kessler, R. C. (2014). Barriers to mental health treatment: Results from the WHO World Mental Health surveys. *Psychological Medicine. 44*(6), 1303-1317.

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: A meta-analysis. *PLoS One*. *5*(10), e13196. doi:10.1371/journal.pone.0013196

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR mental health*, *3*(1), e7.

Barlow, D. H., Ellard, K. K., Fairholme, C. P., Farchione, T. J., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. T. (2011). *Unified protocol for transdiagnostic treatment of emotional disorders: Workbook (treatments that work)*. New York, NY: Oxford University Press.

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *Journal of medical Internet research*, *21*(9), e14567.

Bry, L. J., Chou, T., Miguel, E., & Comer, J. S. (2018). Consumer smartphone apps marketed for child and adolescent anxiety: A systematic review and content analysis. *Behavior Therapy*, *49*(2), 249-261.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7-18.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685-716.

Chorpita, B.F., Daleiden, E., & Weisz, J.R. (2005). Identifying and selecting the common elements of evidence-based interventions: A distillation and matching model. *Mental Health Services Research, 7*(1), 5-20.

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*(3), 566-579.

Coulon, S. M., Monroe, C. M., & West, D. S. (2016). A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. *American Journal of Preventive Medicine*, *51*(1), 95-105.

Craske, M. G. (2010). *Cognitive–behavioral therapy*. American Psychological Association.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*, *76*(6), 909-922.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry*, *58*(7), 376-385.

Dogruel, L., Joeckel, S., & Bowman, N. D. (2015). Choosing the right app: An exploratory perspective on heuristic decision processes for smartphone app selection. *Mobile Media & Communication*, *3*(1), 125-144.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: A systematic review. *Journal of Medical Internet Research. 15*(11), e247.

Ebert, D. D., Zarski, A. C., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M., & Riber, H. (2015). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: A meta-analysis of randomized controlled outcome trials. *PLoS ONE*, *10*(3), e0119895.

Fairburn, C. G., & Patel, V. (2017). The impact of digital technology on psychological treatments and their dissemination. *Behaviour Research and Therapy*, *88*, 19-25.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap ,A., Rosenbaum, S., & Sarris, J. (2017a). The efficacy of smartphone-based mental health interventions for depressive symptoms: A meta-analysis of randomized controlled trials. *World Psychiatry*. *16*(3), 287–298.

Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017b). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *Journal of Affective Disorders*, *218*, 15-22.

Firth, J., & Torous, J. (2015). Smartphone apps for schizophrenia: a systematic review. *JMIR mHealth and uHealth*, *3*(4), e102.

Fleming, T. M., De Beurs, D., Khazaal, Y., Gaggioli, A., Riva, G., Botella, C., ... Merry, S. (2016). Maximizing the impact of e-therapy and serious gaming: Time for a paradigm shift. *Frontiers in Psychiatry*, *7*, 65.

Fleming, T. M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S. N., Cheek, C., ... Hetrick, S. (2017). Serious games and gamification for mental health: current status and promising directions. *Frontiers in Psychiatry*, *7*, 215.

Foa, E. B. (2011). Prolonged exposure therapy: past, present, and future. *Depression and anxiety*.

Frank, R. H., & Cook, P. J. (2010). *The winner-take-all society: Why the few at the top get so much more than the rest of us*. Random House.

Gulliver, A., Griffiths, K., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC Psychiatry*, *10*(1), 113.

Henson, P., David, G., Albright, K., & Torous, J. (2019). Deriving a practical framework for the evaluation of health apps. *The Lancet Digital Health*, *1*(2), e52-e54.

Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child & Adolescent Psychology*, *45*(2), 91-113.

Huguet, A., Rao, S., McGrath, P. J., Wozney, L., Wheaton, M., Conrod, J., & Rozario, S. (2016). A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. *PloS one*, *11*(5), e0154248.

Josephine, K., Josefine, L., Philipp, D., David, E., & Harald, B. (2017). Internet-and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. *Journal of Affective Disorders*, *223*, 28-40.

Juster, H. R., & Heimberg, R. G. (1995). Social phobia: Longitudinal course and long-term outcome of cognitive-behavioral treatment. *Psychiatric Clinics*, *18*(4), 821-842.

Kaplan, R. M., & Stone, A. A. (2013). Bringing the lab and clinic to the community: Mobile technologies for health promotion and disease prevention. *Annual Review of Psychology*, *64*, 471-498.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on psychological science*, *6*(1), 21-37.

Kazdin, A. (2017). Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. *Behaviour Research and Therapy*, *88*, 7-18.

Kendall, P. C., & Hollon, S. D. (Eds.). (2013). *Cognitive-behavioral interventions: Theory, research, and procedures* (Vol. 21). Academic Press.

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry*, *18*(3), 325-336.

Lindsey, M. A., Brandt, N. E., Becker, K. D., Lee, B. R., Barth, R. P., Daleiden, E. L., & Chorpita, B. F. (2014). Identifying the common elements of treatment engagement interventions in children's mental health services. *Clinical Child and Family Psychology Review*, *17*(3), 283-298.

Mani, M., Kavanagh, D. J., Hides, L., & Stoyanov, S. R. (2015). Review and evaluation of mindfulness-based iPhone apps. *JMIR mHealth and uHealth*, *3*(3), e82.

Mohr, D. C., Riper, H., & Schueller, S. M. (2018). A solution-focused research approach to achieve an implementable revolution in digital mental health. *JAMA psychiatry*, *75*(2), 113-114.

Neary, M., & Schueller, S. M. (2018). State of the field of mental health apps. *Cognitive and Behavioral Practice*, *25*(4), 531-537.

Ng, M. M., Firth, J., Minen, M., & Torous, J. (2019). User engagement in mental health apps: a review of measurement, reporting, and validity. *Psychiatric Services*, appi-ps.

Ruscio, A. M. (2019). Normal versus pathological mood: Implications for diagnosis. *Annual review of clinical psychology*, *15*, 179-205.

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth*, *3*(1), e27.

Strosahl, K. D., & Robinson, P. J. (2017). *The mindfulness and acceptance workbook for depression: Using acceptance and commitment therapy to move through depression and create a life worth living*. Oakland, CA: New Harbinger Publications.

Shen, N., Levitan, M. J., Johnson, A., Bender, J. L., Hamilton-Page, M., Jadad, A. A. R., & Wiljer, D. (2015). Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR mHealth and uHealth*, *3*(1), e16.

Van Ameringen, M., Turna, J., Khalesi, Z., Pullia, K., & Patterson, B. (2017). 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. *Depression and anxiety*, *34*(6), 526-539.

Wasil, A.R., Gillespie, S., Shingleton, R., Wilks, C., & Weisz, J. (2020). Examining the reach of smartphone applications for depression and anxiety. *The American Journal of Psychiatry*, *177*(5), 464–465.

Wasil, A.R., Patel, R., Cho, J.Y., Shingleton, R., Weisz, J.R., & DeRubeis, J. Smartphone apps for eating disorders: *A systematic review of evidence-based content and application of user-adjusted analyses*. Manuscript submitted for publication.

Wasil, A. R., Venturo-Conerly, K. E., Shingleton, R. M., & Weisz, J. R. (2019). A review of popular smartphone apps for depression and anxiety: Assessing the inclusion of evidence-based content. *Behaviour Research and Therapy*, *123*, 103498

Wasil, A.R., Weisz, J., & DeRubeis, R. (2020). Three questions to consider before developing a mental health app. *World Psychiatry*, *19*(2), 252–253.

Weisel, K. K., Fuhrmann, L. M., Berking, M., Baumeister, H., Cuijpers, P., & Ebert, D. D. (2019). Standalone smartphone apps for mental health—a systematic review and meta-analysis. *npj Digital Medicine*, *2*(1), 1-10.

Weissman, M. M., Markowitz, J. C., & Klerman, G. L. (2017). *The guide to interpersonal psychotherapy*. New York, NY: Oxford University Press.

Weisz, J.R., Ng, M.Y., & Bearman, S.K. (2014). Odd couple? Re-envisioning the relation between science and practice in the dissemination-implementation era. *Clinical Psychological Science, 2(1),* 58-74. DOI: 10.1177/2167702613501307

Yang, C. H., Maher, J. P., & Conroy, D. E. (2015). Implementation of behavior change techniques in mobile applications for physical activity. *American journal of preventive medicine*, *48*(4), 452-455.

Data Transparency Statement:

A previous manuscript (Wasil et al., 2019) has been published using data from the same dataset. The previous manuscript reported the frequency of treatment elements within mental health apps, but the previous manuscript did not incorporate any usage data. In contrast, this present submission focuses on explaining and applying a novel method of user-adjusted analyses. All of the usage data, and the analyses related to usage data, are original and have not been previously published. We also demonstrate how the results of our user-adjusted analyses differ from the unadjusted analyses in the previous manuscript.