

Running Head: COGNITIVE BEHAVIORAL MOBILE APPLICATIONS

Cognitive behavioral mobile applications: Clinical studies, marketplace overview, and research
agenda

John Torous^{1,2}; Michael Levin³; David Ahern²; Megan Oser²

1 - Harvard Longwood Psychiatry Residency Training Program, Harvard Medical School,
Boston, MA, 02115, USA

2 - Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston,
MA, 02115, USA

3 - Department of Psychology, Utah State University Logan, UT 84322, USA

Please address correspondence to Dr. Megan L. Oser, Department of Psychiatry, 221 Longwood
Ave., 4th Floor, Brigham & Women's Hospital, Boston, MA 02130; Email:
moser@bwh.harvard.edu.

Highlights:

- The literature supports feasibility of cognitive behavioral interventions via smartphones
- A large number of cognitive behavioral apps exist in commercial marketplaces
- There is a need for more clinical evidence before recommending or rating these apps.

Cognitive behavioral mobile applications: clinical studies, marketplace overview, and research agenda

Increasingly, digital technologies are gaining attention as a means to improve access to care and deliver effective therapeutic interventions directly to patients (Luxton, Mccann, Bush, Mishkind, & Reger, 2011; Comer, 2015; Jones et al., 2014; Kazdin, 2015). While computer based therapies have been extensively studied, and overall have been shown effective (Cartreine, Ahern, & Locke, 2010), less is known about smartphone or tablet mobile applications as a tool to deliver evidence based mental health care (Boudreaux et al., 2014; Donker et al., 2013). Smartphone applications offer a novel tool that individuals seeking mental health care now increasingly own and are ready and interested in using for their mental health (Torous et al., 2014). There are many different types and categories of mental health apps and a useful framework is to consider those that seek to help diagnosis and monitor (eg symptom trackers, surveys) versus those that seek to intervene and assist (eg self-help tools, medication reminders, guided therapy). Within either category, apps can have an adjunctive role in clinical care or can also offer stand-alone services without the support and guidance of a clinician. Thus there are a variety of types and uses of mental health apps that are both promising and warrant careful study.

The potential for mobile behavioral health to deliver and increase access to evidence based therapies comes at time of tremendous unmet needs. Anxiety and mood disorders remain the two most prevalent psychiatric illnesses, together impacting billions of people worldwide. Whether their burden is measured in patient suffering, disability, increased mortality, lost productivity, or healthcare costs, the impact is devastating. Although effective treatments exist for both anxiety and mood disorders, limited access to care remains a significant barrier worldwide (Kazdin & Blasé, 2011; Ustün, 1999). But while there is increasing appreciation of

the potential of smartphone applications to deliver care and fill this unmet need; research and clinical experience are still in a nascent stage (Jones et al., 2014; Boudreaux et al., 2014; Powell, Landman, & Bates, 2014). In addition to limited research, other important concerns include privacy and confidentiality of data gathered with CBAs. Emerging research on depression is exploring apps that correlate mood symptoms with GPS location data, revealing a tremendous amount of personal information such as where one works, lives and shops (Saeb et al., 2015). Ensuring that such data is secure and the ethics of collecting such invasive data need to be carefully considered (Kramer, Kinn, & Mishkind, 2015). For example, in October 2015 the British National Health Service closed its online library of curated apps amid strong evidence that approved apps were both not evidence based (Huckvale et al., 2015) and not secure with patient data (Wicks & Chiauzzi, 2015). What one day was a leading example of healthcare app ratings was the next day closed amid public outcry. Further concerns include limited patient adherence and a lack of data on long term patient engagement with apps (Belisario, Kamesk, Huckvale, O'Donoghue, & Morrison, 2015). There is a clear need for further research on many fronts, especially considering how even basic data such as the safety and efficacy of apps is limited. The issue is especially important for mHealth psychotherapy research (Clough & Casey, 2015) and cognitive-behavioral intervention mobile applications (CBAs). Unlike applications that monitor health or suggest healthy lifestyle choices, CBAs have the potential to harness and provide evidence based intervention components that may provide symptom relief and enhance self-management. Conversely, poorly executed, designed, or inadequately implemented applications (i.e., without adequate assessment of consumer understanding or insufficient intensity of intervention) may generate spurious negative results and lead patients and providers to believe that a particular treatment approach has failed when such is not true. This missed

opportunity due to low quality or inadequately implemented apps may reduce future adoption of mHealth and possibly reduce patient confidence/ engagement in cognitive-behavioral interventions more generally. Both providers and patients need empirically derived guidelines to evaluate the quality of the multitude of available apps while identifying those that may be ineffective.

In this article we review the literature regarding the feasibility and efficacy of smartphone based CBAs, briefly examine trends in the application marketplace and how this impacts traditional research endeavors. We conclude with putting forth a clinical research agenda bridging the need for CBA evaluation in the context of the industry-led marketplace as next steps for the applied psychiatric and behavioral intervention research field.

Method

To examine the CBA evidence base, we searched for publications on behavioral activation (BA), CBT, dialectical behavior therapy (DBT), or Acceptance and Commitment Therapy (ACT) (traditional and contextual CBT (Hayes, Villatte, Levin, & Hildenbrandt, 2011)) delivered via smartphone. Published studies were identified through searches conducted on PsycInfo and Medline using keywords terms “mobile app” and “mobile” and “smartphone” in pairwise combination (“AND”) with the following terms including “Cognitive Behavioral Therapy”, “Behavioral Activation”, “Dialectical Behavior Therapy”, and “Acceptance and Commitment Therapy.” We searched for papers published before February 2015, without any limitations on the initial date of publication. Inclusion criteria were: [1] must present quantitative results on clinical outcomes, [2] feature a modern smartphone defined as a phone with internet connectivity, the ability to download applications, and either a touch screen or ‘qwerty’ keyboard, and finally [3] evaluate one of the four cognitive behavioral therapeutic approaches listed above

for at least a one week duration. By specifying that quantitative results on clinical outcomes are required, we selected studies with at least one numerical measure of clinical outcomes (e.g., PHQ-9 or single item rating of anxiety/distress) which is in contrast to results such as satisfaction, usability, and interest which are important but not the topic of this paper. Of note, many apps may be inspired by or feature elements of various CBTs, but our goal here was to only evaluate those papers focused on the evaluation of CBT approaches (i.e., translations of existing CBT methods to a mobile app format). However, we did not exclude research studies that tested more limited dosages or components of mobile apps (e.g., an app study only evaluating one component of CBT could still be included). A one-week minimum duration was chosen to provide a minimum window representative of clinical evaluation of a mobile app, as opposed to more constrained use that is unlikely to impact outcomes and more representative of usability testing/formative evaluation. Follow up papers reporting feature analyses or qualitative results of app studies meeting the above inclusion criteria were also included, provided they were connected to a published outcome study. We did not include a criterion for design rigor given that we expected much of the literature to be initial pilot research, which we wanted to ensure was included. Papers that focused on text-messaging interventions or utilized customized hardware or devices were excluded. These criteria excluded many well-known and deployed CBAs such as PTSD Coach (Kuhn et al., 2014) which highlights the challenges in matching the research literature to the successful and timely deployment of mobile apps.

Results

The literature review identified a total of 9 studies which met the above criteria, one study evaluating a BA app, three on CBT, one on DBT, and four on ACT (see Table 1). Of these, four were randomized controlled trials (RCTs) and the others were mainly pilot, feasibility, or

secondary analysis studies. The mean duration of participant app use was 5.4 weeks, the minimum duration of app use was 2 weeks and maximum duration of use was 9 weeks. Four of the nine clinical studies included a follow up period and the mean duration of follow up was 3.5 months. Five studies used clinical populations (participants with depression, social anxiety, substance use disorder, etc) and four studies relied on non-clinical populations.

CBT Studies

Three studies were found testing traditional cognitive -behavioral apps (i.e., not BA, DBT or ACT), which generally provided initial support for their use. One RCT (Watts, Mackenzie, Thomas, & Griskaitis, 2013) included 35 participants with major depressive disorder who were randomized to computer based CBT or smartphone based CBT lessons for eight weeks. In this study, the ‘Get Happy’ app offered CBT lessons that read like a comic book and encouraged participants to complete homework and relevant activities. Approximately 69% of participants in the app group completed all six lessons. Over six weeks, the CBA significantly improved and lowered PHQ-9 pre to post depression scores (Cohen’s *d* value 1.41) although effects were similar, and not statistically different between the computer CBT groups.

Another RCT examined the efficacy of a CBA to treat social anxiety in adults (Dagöo et al., 2014). In this study, 52 subjects were randomized to use a CBA or an interpersonal psychotherapy-based app for nine weeks. The CBA consisted of nine weekly modules that were largely text based, although included video feedback features as well as homework that could be submitted electronically. Of the 30 subjects randomized to the CBT app, 63% completed all treatment modules and, in the interpersonal psychotherapy-based app condition, 52% completed all modules. Results indicated that while both groups showed improvement in symptoms, the

CBT app performed better on self-reported social anxiety, as measured by the LSAS-SR, with a pre to post Cohen's d value of .99 and three month follow up Cohen's d value of .93.

A third within-subject pilot demonstrated feasibility of delivering CBT exercises for healthy adults with stress via a smartphone application (Morris et al., 2010). Ten participants used the app for one month. The app allowed participants to indicate and rate the intensity of their emotions and then offered abbreviated CBT interventions such as breathing visualization, guided relaxation, and reappraisal exercises. Results were presented as interview narratives and experience sampling data gathered from the app. The study demonstrated the ability of an app to increase self-awareness and coping skills; and is also notable for being the first study of its kind to use smartphone technology.

BA Studies

Only one study was found testing a BA app. A RCT (Ly et al., 2014a) indicated the feasibility of BA delivered over a custom smartphone application as compared to a mindfulness app with a sample of 81 depressed participants. Significant within condition reductions were found for BA on two self-reported depressive symptom measures (PHQ-9 and BDI-II) at post-assessment and sustained at six month follow-up. Within condition effect sizes were large (Cohen's d values ranging 1.14 -1.83) for the BA app condition, but were equivalent to effects found with the mindfulness app condition. A second paper reported on participants qualitative experiences using the app, which were positive. This paper also suggested that therapist involvement might be an important factor influencing patients' experiences and outcomes with apps (Ly et al., 2015).

DBT Studies

A within-subject pilot study demonstrated the feasibility of a DBT smartphone app for delivering the opposite action skill in patients with borderline personality disorder and comorbid substance use as an adjunct to face-to-face DBT (Rizvi, Dimeff, Skutch, Carroll, & Linehan, 2011). Twenty-two participants used the 'DBT Coach' app which assessed emotional intensity, asked participants to identify their emotions, and then guided them to apply the opposite action skill. On average, participants used the app 15 times over the approximately two week study period. Eighty-five percent of participants were adherent to daily assessments via the app and overall satisfaction scores were high. Within condition effects indicated significant reductions in emotional intensity (Cohen's $d = .52$) and urges to use substances (Cohen's $d = .29$) after completing each coaching session. Furthermore, there were significant improvements from pre to post app testing on depression (BDI; Cohen's $d = .55$) and other psychiatric symptoms (BSI; Cohen's $d = .43$).

ACT Studies

Four ACT mobile app studies were conducted and provide support for the use of ACT delivered via a smartphone. A RCT of a novel smoking cessation app, 'SmartQuit' was conducted with 196 participants who were randomized to the study app or the National Cancer Institute's smoking cessation app called 'QuitGuide.' The 'SmartQuit' app was designed as a self paced ACT intervention based on ACT's theory-based process of change. Ninety-eight participants were randomized to the ACT app and 84% of these completed the eight week study. Fifty three percent of participants felt that the ACT app was useful for quitting. Quit rates for those using the ACT app were 13%, and for the National Cancer Institute's smoking cessation app were 8% (Odds Ratio = 2.7) , although there was not a statistically significant difference for quit rates between both apps (Bricker et al., 2014). Interestingly, a feature analysis on the same

ACT app study data noted the features participants most utilized were not those most correlated with predicting smoking abstinence, suggesting the complexity of app use and need for further study (Heffner, Vilaradaga, Mercer, Kientz, & Bricker, 2015).

Another RCT of an ACT based smartphone app examined the impact of an app intervention on stress in a sample of 73 middle managers (Ly, Asplund, & Andersson, 2014b). The six week study randomized participants to a waitlist control or to the app group, which consisted of six modules providing education on ACT principles to handle stress. Of the 36 participants randomized to the app, 44% were adherent to the app over the entire six weeks. The ACT app group showed improved outcomes from pre to post on stress/distress outcomes including the GHQ-12 ($d = .37$) and PSS-14 ($d = .62$) and significantly greater effects relative to the waitlist condition.

One of the earliest ACT smartphone studies investigated the feasibility of delivering ACT via a smartphone (Ly, Dahl, Carlbring, & Andersson, 2012). The study involved 11 non-clinical participants who used an ACT app called Viary for over a 4 week period. The Viary app allowed participants to define, remember, and register behaviors in line with their values. Although primarily an exploratory study, results demonstrated pre to post improvements on values $d = .77$ (EVS) and psychological flexibility $d = .50$ (AAQ-II) although no significant improvements on psychological symptoms or life satisfaction.

Another feasibility study tested an app called 'Ovia' which was designed to support active learning of ACT skills in daily life (Ahtinen et al., 2013). Fifteen non-clinical participants were provided with the app for a one month period, and used it on average for 11.5 days during the study period. Results showed a pre to post improvements on stress ($d = 3.8$) and life satisfaction ($d = 2.6$) although no improvements on psychological flexibility (AAQ).

Overall, while these papers are notable for pioneering systematic investigations of CBAs, it would be premature to base practice recommendations on such a small amount of diverse pilot data with varying degrees of methodological rigor. Similarly, although it is difficult to conclude the mean duration of app use necessary for participants to attain new cognitive behavioral skills, given the diversity of apps and outcomes in the reviewed studies, it appears that after two weeks of use in conjunction with individual therapy, positive effects from app interventions may be observed.

‘INSERT TABLE 1 ABOUT HERE’

Brief Overview of the Marketplace

Providers have a small evidence base to turn to when considering apps for their patients. Conversely, patients likely face the opposite situation in that they are presented with a wide array of applications on the public marketplaces of the Apple App or Google Play stores. We searched the Apple App Store and Google Play Store on March 1st, 2015 for the search terms listed below and recorded the number of hits. While there are numerous potential search terms and means to search these commercial marketplaces, the goal here is to highlight broad trends rather than quantify the exact number of applications. It is difficult to capture the exact number of applications available in the commercial space as different search engines and platforms yield different results and applications are frequently added and removed from the marketplace. Also, we did not control for the fact that some apps may be listed twice, once on each app store. Unlike academic search engines, which are designed to archive literature, searching commercial markets is more dynamic given the intended audience is consumers. The results of this search may be seen in table 2.

We also searched whether the apps featured in the reviewed studies were available to download on the Apple App or Google Play stores. Two of the four ACT apps that have published outcome findings, Oiva (Ahtinen et al., 2013) and SmartQuit (Bricker et al., 2014) were available for download (on Android, and Android and Apple respectively). The DBT app, DBT coach, was not available for download. The three papers by Ly and colleagues (2012; 2014a; 2014b) featured evolving or modified versions of a BA/ACT app called Viary, which is available to download on Apple and Android for those with a username and password provided as part of a professional system. There was no evidence that any other apps featured in the reviewed papers were available. It is also important to note that of those studied apps we did find, it is difficult to verify which aspects of the app available today are the same as those described in the clinical studies as apps often change with new versions. Thus, of the 9 CBAs that have published outcome research, only three (Oiva, SmartQuit, Viary) appear to be commercially available to end users in some format.

‘INSERT TABLE 2 ABOUT HERE’

Discussion

Our study revealed 9 clinical studies and a total of 11 papers with evidence for smartphone delivered cognitive behavioral interventions and 322 iOS and 125 Android related applications available for immediate use by patients on the commercial marketplaces. Only 33% of these applications reviewed in published studies are available for download and use on the marketplace (e.g., Oiva, SmartQuit, Viary), making it generally difficult for a provider to make direct recommendation based solely on the scientific literature. There is currently not enough

empirical literature to provide evidence-based guidelines for selecting and implementing CBAs. Given the state of the literature, rather than proposing practice guidelines for using apps, we will suggest an agenda for how to move research forward in order to inform such practice guidelines.

Mobile apps present a number of unique challenges (Kumar et al., 2013; Mohr, Cheung, Schueller, Hendricks Brown, & Duan, 2013) which may help account for the discrepancy we found between the prevalence of deployed CBAs in the marketplace relative to the published outcome research. The mobile app development/ deployment cycle is much faster than those of traditional face-to-face therapy protocols, which is necessary given the rapid progress in mobile technology as well as the competitiveness of the mobile app marketplace (Boudreaux et al., 2014). This environment stands in marked contrast to the standard phased research development and testing cycle previously adopted by National Institutes of Health, in which RCTs are the primary evaluation method despite typically taking 5.5 years on average to complete (Ioannidis, 1998). Requiring mobile apps to be subjected to the same standard RCT methods prior to deployment is not feasible as the technology is very likely to become outdated and the commercial niche very likely to have been addressed through other less “research conscientious” development teams within that time span. Moreover, the traditional RCT with intact groups may not be appropriate for evaluating highly individualized interventions intrinsic to mobile apps that can be tailored to specific individual characteristics and attributes. Mobile apps must be studied rigorously, but methods beyond RCTs may also be appropriate and necessary.

Steps Forward: CBA Research Agenda

Well-designed and controlled studies testing the efficacy of specific apps with clinical populations are still clearly needed as demonstrated by the predominantly feasibility and small-scale studies identified in our review of the literature. A research agenda tailored specifically for

the unique challenges of mobile apps, and especially CBAs is needed, some possible features of which we outline below.

Focusing on evidence-based criteria for rating cognitive-behavioral mobile apps.

One potential strategy is to focus on developing a standardized set of evidence-based rating criteria for mobile apps. Such criteria may identify which mobile apps are more (or less) likely to be efficacious, safe, and clinically indicated for use providing guidance to practitioners, patients, and other stakeholders while more time consuming RCTs are conducted. Further, these guidelines are needed even when relevant RCTs are made available since such studies will never provide direct evidence for every mobile app on the market. Rather some degree of generalization from the tested mobile app to similar mobile apps is necessary; similar to generalization of findings from an outcome study in one clinic to other clinics that implement the same treatment. A set of evidence-based criteria could provide a useful guide for making such generalizations (i.e., how similar is this app to other efficacious app on key criteria?), reducing the immediate burden on conducting RCTs with each app before it can be recommended.

Research on rating smartphone apps is still nascent. While the evidence may be modest, examples of mobile app rating criteria exist, often focusing on whether mobile apps include intervention strategies known to be efficacious in other mediums (i.e., face-to-face interventions) (Abroms, Westmaas, Bontemps-Jones, Ramani, & Mellerson, 2013; Pagoto, Schneider, Jojic, DeBiasse, & Mann, 2013). Broad frameworks for evaluating mobile mental health apps includes consideration of usefulness, usability, integration, and infrastructure (Chan, Torous, Hinton, & Yellowlees, 2015) and consideration of all ASPECTs on a mental health app in its evaluation: Actionable data, Secure use, meeting Professional standards, Evidence based, Customizable for patient needs, and Transparent data policies (Torous, Chan, Yellowlees, & Boland, 2016). The

UK's British Standards Institute released a code of practice proposal which sets forth highly specified best practice principles for healthcare related apps although is technically not a rating system itself ("Code of Practice", 2015). Others have proposed more specific rating tools, with the mobile app rating scale (MARS) (Stoyanov et al., 2015) which is a 23 item scale that claims strong internal consistency and inter-rater reliability. The difficulty of implementing app rating for mental health was recently underscored in a study that examined 22 general criteria for app ratings (eg ease of use, perceived effectiveness, basis of research, password protection etc) and found wide variation in the inter-rater reliability between six evaluators with healthcare backgrounds (Powell et al., in press). Advances in app research will directly translate into advances in rating apps, but at this point there is no leading evidence based rating system or tool.

Specific to CBAs, several organizations provide ratings based often on more rationally-derived criteria that might be a priority to stakeholders such as usability and functionality and clinical judgment regarding potential safety and efficacy (e.g., Anxiety and Depression Association of America (ADAA) <http://www.adaa.org/finding-help/mobile-apps>; Association for Cognitive and Behavioral Therapies <http://www.abct.org/Resources>). Another potentially useful model, such as that presented by Boudreax and colleagues, is general mobile app websites (clearinghouses) in which the goal is to evaluate the usability, functionality, content accuracy, and evidence base for the app (Boudreax et al., 2013) One example is the nonprofit organization Psyberguide (www.psyberguide.org). However in light of the safety and efficacy concerns of apps that were on the now defunct and closed National Health Service app rating site (Torous, in press) caution must be taken when using external evaluation tools. Although ratings on usability, security, privacy, and similar features can likely be made by trained reviewers to address important consumer issues, such criteria do not answer the key question of whether the mobile

app is likely to be efficacious. Even criteria rating whether apps use evidence-based strategies is limited as this assumes such methods known to be efficacious in other modalities are applicable and efficacious in a mobile app format (Abroms et al., 2013).

However, it is also possible that certain evidence-based strategies could be iatrogenic (at least if not implemented adequately) in a mobile app format (e.g., providing a low dosage/weakly structured exposure intervention that leads to treatment failure and reinforces avoidance). Thus, research is also needed to identify qualities of mobile apps which might lead to iatrogenic effects (e.g., suggesting maladaptive coping strategies that increase emotional dysregulation, inadequate supports for behavioral methods like exposure or goal setting that increase distress and reduce engagement in treatment). Furthermore, identifying strategies that are simply inert and unlikely to produce gains is no less important as this may lead to patients disengaging from treatment seeking and a missed opportunity to provide more helpful services. As outcome research on CBAs grows, the development of a more empirically derived evaluation criteria could be key for informing the public of the likely efficacy of such programs. Although these review criteria are in early stages of development, there are some common themes for clinicians looking to review and select mobile apps now. For example, common factors identified as important include ease of use, appropriate handling of data for privacy/confidentiality, and potential efficacy.

Examining components and mechanism of change in mobile apps. One potential area for informing evidence-based criteria is identifying the cognitive behavioral mechanisms of change and treatment components that can be effectively targeted in mobile apps to produce meaningful outcomes (e.g., skills generalization, self monitoring, cognitive restructuring or defusion, goal setting). The importance of this was demonstrated in feature analysis of the Smart

Quit smoking cessation app which demonstrated that only a minority of the most popular ACT and CBT specific features of the app were actually associated with smoking abstinence (Heffner et al., 2015). Qualitative analysis of another ACT app suggested that therapist involvement may also be a critical component and has a motivating function (Ly et al., 2015). Conducting component tests and process of change analyses will identify components and mechanisms accounting for positive effects and potential negative effects. Similarly, sophisticated methods such as Sequential Multiple Assignment Randomized Trial (SMART; (Collins, Murphy, & Stretcher, 2007)) could guide rapid, systematic testing of various features and components. A list of such mechanisms and components could be used to identify what strategies delivered via CBAs are likely efficacious.

Considering a broader range of outcomes. Another need is to clarify and expand the set of outcomes that define efficacy for CBAs. Symptom reduction and improvements in behavioral functioning are often specific targets with apps (as seen in 8 out of 11 studies included in our review (Ly et al., 2014a, 2014b; Watts et al., 2013; Dagöo et al., 2014; Rizvi et al., 2011; Bricker et al., 2014; Ly et al., 2012; Ahtinen et al., 2013)). There are other outcomes to consider that are relevant to consumers, therapists and other stakeholders, some of which are more clearly linked to specific app rating criteria. As such, we recommend that CBA research also evaluate the following: improving the efficiency of face-to-face treatment, maintaining engagement/retention in therapy, relapse prevention and maintenance of treatment gains, improvements in therapeutic mechanisms of change, skills acquisition/generalization, and patient and therapist satisfaction. For example, research might elucidate the mobile app features most likely to promote relapse prevention (e.g., skills generalization, development of coping plans for

high risk situations) or increase patients' continued engagement in therapy (e.g., homework tools, therapist communication features, self-monitoring).

Understanding the primary purpose (improve one skill vs. multi-component intervention) and target (self-help only vs adjunct to therapy) of CBAs. While CBAs may have the common goal of improving the users' condition, they may function differently to serve as stand-alone interventions vs adjunctive vs non-clinical self-help tools. Understanding the particular niche of a given CBA is a complex issue although several studies (Ly et al. 2014a; Rizvi et al., 2011) begin to shed light on some pertinent questions. For instance, one study (Ly et al., 2014a) found that a BA mobile app may be more effective for those with severe depression and a mindfulness mobile app may be more effective for milder depression. Rizvi et al., 2011 tested one component of DBT emotion regulation skills in conjunction with individual in person DBT. Furthermore, another study (Watts et al., 2013) tested whether a mobile app was comparable in effectiveness to computer delivered CBT. Increased focus on these outcomes and implementation methods can further highlight criteria for mobile apps that are most likely to be beneficial for patients, therapists, providers, and larger systems of care.

Expanding the methods used to test apps. Some of the above criteria will likely be identified through RCTs; however, this work would be significantly bolstered by using a more diverse range of methods (Kumar et al., 2013; Mohr et al., 2013). A more "agile science" approach that is iterative and individualized may be preferred (www.agilscience.org). Agile science refers to a novel scientific approach that promotes a rapid, iterative/evaluative process to improve efficiencies and to address complex societal problems. This approach is particularly suited to the evaluation of mobile apps designed to deliver behavior change interventions since it

incorporates the *agile* method of development from the tech sector and favors rapid prototyping and pragmatic scientific rigor.

One notable example is the Continuous Evaluation of Evolving Behavioral Intervention Technologies method (CEEBIT; (Mohr et al., 2013)), which involves conducting ongoing evaluation of a deployed technology and, as updated versions are released, comparing these versions to determine relative efficacy. This approach not only provides the opportunity to evaluate mobile apps without delaying deployment, but also allows the identification of how various versions affect efficacy through continuous evaluation (which could inform establishment of rating criteria if version updates are well specified). Of note, none of the 9 papers we identified used the CEEBIT method. Instead, all studies were conducted on a single version of the app for a predetermined and discrete amount of time.

A related set of methods are those involving intensive longitudinal data, which is often made available within mobile apps and can similarly be collected during deployment. The data density and opportunities to examine it in relation to program usage and other contextual variables can similarly lead to the identification of important features, components, and mechanisms. This dense set of data also allows for potentially stronger inferences regarding efficacy based on single subject designs. For example, although none of the CBAs presented in this review used this methodology, recent studies of non-CBA apps for smoking cessation and alcohol abuse collected numerous sources of data including, photographs (Hertzberg et al., 2013), location (Gustafson et al., 2014), and exact times of app usage (Watkins et al., 2014). Given the technology to capture and explore such dense data exists, CBAs may also benefit from trialing such methodology in future studies.

Selection of measures. In reviewing the literature, it became apparent that measurement selection needs to be refined and researchers cannot simply map on measures used in CBT psychotherapy outcome research. The heterogeneity of measurement tools makes comparative analysis of different studies, even between the same app, difficult. Many CBA studies identified in this article reported data on participant adherence to the application (Ly et al., 2014a; 2013; Rizvi et al., 2011) but it is unknown exactly how rates of adherence to a smartphone application translate into adherence with treatment and the degree of adherence necessary to achieve benefits. Perhaps new mobile-app based measures analogous to CBT outcome scales could be developed and validated. Furthermore, choice of scales needs to be theoretically and empirically tied to extant literature with particular attention paid to the assessment schedule and duration of intervention.

Immediate strategies for selecting apps.

While commercial marketplaces, such as Apple app and Google Play stores, contain numerous apps, our results have shown that few apps that are clinically studied are actually available for download and thus it is difficult to use the research base as a guide. The closing of the NHS App Library underscores the difficulty of rating apps in a time where we are still lacking the research to determine CBA gold standards including the rating criteria. Given that the above research agenda will take time to implement and yet patients and clinicians are encountering mobile apps currently, we also want to briefly discuss strategies that providers can employ today in app selection

When considering apps within commercial marketplaces:

- Identify apps that clearly indicate how patient data is protected and kept confidential. If an app does not make this clear, it is possible that it is selling patient data or exposing personal information for the world to see.
- While not a guarantee of quality, apps with university affiliations may be helpful as well as across checking for inclusion in the new and evolving cognitive behavioral databases and resources – ADAA’s <http://www.adaa.org/finding-help/mobile-apps> and ABCT: <http://www.abct.org/Resources>.
- Clinicians may find utility in a quick ‘eyeball test’ of downloading the app and using it themselves to at least ensure it does not provide contradictory information to establish standards of care.
- Consult with other providers, either directly in the clinic or at local meetings, digitally through professional listservs, or facilitated through professional organization beginning to compile lists of recommended programs (e.g., http://contextualscience.org/act_and_other_cbsrelevant_technologies).
- Health insurance companies offer another source for vetted apps, such as Aetna, which features downloadable apps on their behavioral and mental health program webpages.

As demonstrated in our paper, seeking evidence from clinical studies of CBA apps is unlikely to be fruitful given the paucity of research and plethora of commercial apps. Given the many uncertainties with using apps at this time, it is advisable to document a conversation with the patient beforehand about these numerous unknowns and potential risks. Evaluating the risk

benefit ratio for each patient and exercising prudent clinical judgement may be the most effective tool currently available for picking these apps.

Conclusions

The published outcome literature on CBAs is currently small, yet commercial marketplace data suggests a strong and flourishing interest in CBAs. While the majority of the eleven CBA studies identified in this review were feasibility/pilot studies, it is noteworthy that all showed positive or promising results. Although evaluating CBAs based on traditional research methodologies (e.g., RCTs) may not be optimal given the rapid development of numerous new CBAs and the lengthy timeline for journal publication, focusing on evidence based rating criteria and exploring new research methodologies best suited for CBAs will ensure that our field retains an active and critical role in this new digital mental health frontier. As the empirically-derived literature on CBAs continues to expand, the importance of such emerging methodologies will quickly be realized.

References

- Abroms, L. C., Lee Westmaas, J., Bontemps-Jones, J., Ramani, R., & Mellerson, J. (2013). A content analysis of popular smartphone apps for smoking cessation. *American Journal of Preventive Medicine, 45*(6), 732–736.
- Ahtinen, A., Mattila, E., Väلكkynen, P., Kaipainen, K., Vanhala, T., Ermes, M., . . . Lappalainen, R. (2013). Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. *Journal of Medical Internet Research mHealth and uHealth, 1*(2), e11.
- Belisario, J.S, Kamesk, J., Huckvale, K., O'Donoghue, J., & Morrison, C.P. (2015). Can apps be used for the delivery of survey questions in public health and clinical research? *Cochrane Database of Systematic Reviews, 2015*(7), doi:10.1002/14651858.MR000042.pub2
- Boudreaux, E. D., Waring, M. E., Hayes, R. B., Sadasivam, R. S., Mullen, S., & Pagoto, S. (2014). Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *Translational Behavioral Medicine, 4*(4), 363-371
- Bricker, J. B., Mull, K. E., Kientz, J. A., Vilardaga, R., Mercer, L. D., Akioka, K. J., & Heffner, J. L. (2014). Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. *Drug and Alcohol Dependence, 143*, 87–94.
- Bush, N. E., Skopp, N., Smolenski, D., Crumpton, R., & Fairall, J. (2013). Behavioral screening measures delivered with a smartphone app: psychometric properties and user preference. *The Journal of Nervous and Mental Disease, 201*(11), 991–995.
- Cartreine, J. A., Ahern, D. K., & Locke, S. E. (2010). A roadmap to computer-based psychotherapy in the United States. *Harvard Review of Psychiatry, 18*(2), 80–95.

Chan, S., Torous, J., Hinton, L., & Yellowlees, P. (2015). Towards a Framework for Evaluating Mobile Mental Health Apps. *Telemedicine and E-Health*, 21(12),1038-1041.

Clough, B. A., & Casey, L. M. (2015). Therapy on the Move.: *International Journal of Cyber Behavior, Psychology and Learning*, 5(1), 33-41.

Code of Practice for wellness apps published. (2015, May 5). Retrieved from

<http://www.bsigroup.com/en-GB/about-bsi/media-centre/press-releases/2015/may/Code-of-Practice-for-wellness-apps-published/#.VrFShrIrIQh>

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART). *American Journal of Preventive Medicine*, 32(5), S112–S118.

Comer, J. S. (2015). Introduction to the special series: applying new technologies to extend the scope and accessibility of mental health care. *Cognitive and Behavioral Practice*. Advance online publication. [doi:10.1016/j.cbpra.2015.04.002](https://doi.org/10.1016/j.cbpra.2015.04.002)

Dagöo, J., Asplund, R. P., Bsenko, H. A., Hjerling, S., Holmberg, A., Westh, S., . . . Andersson, G. (2014). Cognitive behavior therapy versus interpersonal psychotherapy for social anxiety disorder delivered via smartphone and computer: a randomized controlled trial. *Journal of Anxiety Disorders*, 28(4), 410–417.

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.

Gustafson, D. H., McTavish, F. M., Chih, M.-Y., Atwood, A. K., Johnson, R. A., Boyle, M.

G., . . . Shah, D. (2014). A smartphone application to support recovery from alcoholism: a

randomized clinical trial. *Journal of the American Medical Association Psychiatry*, 71(5), 566-572.

Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology*, 7, 141-167

Heffner, J. L., Vilardaga, R., Mercer, L. D., Kientz, J. A., & Bricker, J. B. (2015). Feature-level analysis of a novel smartphone application for smoking cessation. *The American Journal of Drug and Alcohol Abuse*, 41(1), 68–73.

Hertzberg, J. S., Carpenter, V. L., Kirby, A. C., Calhoun, P. S., Moore, S. D., Dennis, M. F., . . . Beckham, J. C. (2013). Mobile contingency management as an adjunctive smoking cessation treatment for smokers with posttraumatic stress disorder. *Nicotine & Tobacco Research : Official Journal of the Society for Research on Nicotine and Tobacco*, 15(11), 1934–1938.

Huckvale, K., Prieto, J. T., Tilney, M., Benghozi, P., & Car, J. (2015). Unaddressed privacy risks in accredited health and wellness apps: A cross-sectional systematic assessment. *BioMed Central Medicine*, 13(1).

Ioannidis, J. P. A. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Journal of the American Medical Association*, 279(4), 281–286.

Jones, D. J., Anton, M., Gonzalez, M., Honeycutt, A., Khavjou, O., Forehand, R., & Parent, J. (2014). Incorporating mobile phone technologies to expand evidence-based care. *Cognitive and Behavioral Practice*. Advance online publication. DOI: 10.1016/j.cbpra.2014.06.002

- Kazdin, A. E. (2015). Technology-based interventions and reducing the burdens of mental illness: perspectives and comments on the special series. *Cognitive and Behavioral Practice*. Advance online publication. DOI: 10.1016/j.cbpra.2014.06.002
- 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.
- Kramer, G. M., Kinn, J. T., & Mishkind, M. C. (2015). Legal, Regulatory, and Risk Management Issues in the Use of Technology to Deliver Mental Health Care. *Cognitive and Behavioral Practice*, 22(3), 258-268.
- Kuhn, E., Greene, C., Hoffman, J., Nguyen, T., Wald, L., Schmidt, J., ... & Ruzek, J. (2014). Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. *Military Medicine*, 179(1), 12-18.
- Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., . . . Swendeman, D. (2013). Mobile health technology evaluation. *American Journal of Preventive Medicine*, 45(2), 228–236.
- Ly, K. H., Dahl, J., Carlbring, P., & Andersson, G. (2012). Development and initial evaluation of a smartphone application based on acceptance and commitment therapy. *SpringerPlus*, 1(11).
- Ly, K. H., Truschel, A., Jarl, L., Magnusson, S., Windahl, T., Johansson, R., . . . Andersson, G. (2014a). Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. *British Medical Journal Open*, 4(1), e003440–e003440.
- Ly, K. H., Asplund, K., & Andersson, G. (2014b). Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. *Internet Interventions*, 1(3), 95–101.

- Ly, K. H., Janni, E., Wrede, R., Sedem, M., Donker, T., Carlbring, P., & Andersson, G. (2015). Experiences of a guided smartphone-based behavioral activation therapy for depression: a qualitative study. *Internet Interventions*, 2(1), 60–68.
- Luxton, D. D., Mccann, R. A., Bush, N. E., Mishkind, M. C., & Reger, G. M. (2011). MHealth for mental health: Integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice*, 42(6), 505-512.
- Mohr, D. C., Cheung, K., Schueller, S. M., Hendricks Brown, C., & Duan, N. (2013). Continuous evaluation of evolving behavioral intervention technologies. *American Journal of Preventive Medicine*, 45(4), 517–523.
- Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M., & Deleeuw, W. (2010). Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. *Journal of Medical Internet Research*, 12(2), e10.
- Pagoto, S., Schneider, K., Jojic, M., DeBiasse, M., & Mann, D. (2013). Evidence-based strategies in weight-loss mobile apps. *American Journal of Preventive Medicine*, 45(5), 576–582.
- Powell, A. C., Landman, A. B., & Bates, D. W. (2014). In search of a few good apps. *Journal of the American Medical Association*, 311(18), 1851–1852.
- Powell, A.C, Torous, J., Chan, S., Raynor, G.S., Shwartz, E., Landman, A.B, (in press). *Inter-rater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Apps. Journal of Medical Internet Research Mobile Health and Ubiquitous Health.* (in press).

- Rizvi, S. L., Dimeff, L. A., Skutch, J., Carroll, D., & Linehan, M. M. (2011). A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. *Behavior Therapy, 42*(4), 589–600.
- Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of Medical Internet Research, 17*(7).
- 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. *Journal of Medical Internet Research Mobile Health and Ubiquitous Health, 3*(1).
- Torous, J., Chan, S. R., Yee-Marie Tan, S., Behrens, J., Mathew, I., Conrad, E. . . . Keshavan, M. (2014). Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. *Journal of Medical Internet Research Mental Health, 1*(1), e5.
- Torous, J. (in press). Mobile telephone apps in mental health practice: uses, opportunities and challenges. *British Journal of Psychiatry Bulletin*.
- Torous J., Chan C., Yellowlees P., & Boland R., 2016. To Use or Not? Evaluating ‘ASPECTS’ of smartphone apps and mobile technology for clinical care in psychiatry. Manuscript submitted for publication.
- Ustün, T. B. (1999). The global burden of mental disorders. *American Journal of Public Health, 89*(9), 1315.
- Watkins, K. L., Regan, S. D., Nguyen, N., Businelle, M. S., Kendzor, D. E., Lam, C . . . Reitzel, L. R. (2014). Advancing cessation research by integrating EMA and geospatial

methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. *Nicotine & Tobacco Research : Official Journal of the Society for Research on Nicotine and Tobacco*, 16 Suppl 2, S93–101.

Watts, S., Mackenzie, A., Thomas, C., & Griskaitis, A. (2013). CBT for depression: a pilot RCT comparing mobile phone vs. computer. *BioMed Central Psychiatry*, 13(49).

Wicks, P., & Chiauuzzi, E. (2015). ‘Trust but verify’ – five approaches to ensure safe medical apps. *BioMed Central Medicine*, 13(1).

Table 1. Summary of published research on CBAs.

Author	Problem area / Interventions	Sample and Study Groups	Research Design	App Time / Frequency of Use	Level of Provider Involvement	Acceptability/Program Usage	Primary outcomes (measurement scales/tools) and within condition effect sizes*
<i>Behavioral Activation</i>							
Ly et al., 2014	Depression, Psychoeducation, activity scheduling, monitoring tool	Depressed sample with BA $n = 40$ Mindful $n = 41$	RCT (BA vs. Mindful app)	8 week / User-initiated	Maximum of 20 minutes per week therapist contact and therapist could send text messages to users	63% adhered to using the BA app all 8 weeks.	No difference between BA vs. Mindfulness. BA improved on depression: pre to post $d = 1.63$ (PHQ-9) and $d = 1.83$ (BDI-II); pre to 6 month follow up $d = 1.14$ (PHQ-9) and $d = 1.19$ (BDI-II).
<i>Cognitive Behavioral Therapy</i>							
Watts et al., 2013	Depression / CBT self-help sessions and homework	Depressed sample with CBT app $n = 15$ cCBT $n = 20$	RCT (CBT app vs. cCBT)	8 week / Six user initiated	Email and phone calls from clinicians for first 1/3 of CBT lessons	69% completed all six app lessons (across conditions). 54% very satisfied with app and 64% very confident to recommend to friend	No difference in depression (PHQ-9) between CBT app and cCBT. CBT app improved on depression (PHQ-9): pre to post $d = 1.41$.
Morris et al., 2010	Emotional self-awareness / Self-reflection and coping skills	8 with high self-rating of stress	Non-randomized	4 week / EMI multiple times daily + user initiated	Weekly meetings with a clinical psychologist.	5 case studies of using the app to increase self-awareness and cope with stress	Feasibility/Acceptability (satisfaction, feature use, and self-reported data on emotional states)
Dagoo et al., 2014	Social Anxiety / Cognitive interventions, exposure, maintenance	SAD sample with CBT app $n = 27$ IPT app $n = 25$	RCT (CBT vs. IPT app)	9 week / User initiated sessions	Weekly therapist check in	63% completed all modules of the CBT app.	Greater pre to post improvements on social anxiety (LSAS) with ACT app vs. IPT app. CBT app improved on social anxiety (LSAS): pre to post $d = .99$, pre to 3 month follow up $d = .93$.
<i>Dialectical Behavior Therapy</i>							
Rizvi et al., 2010	BPD and comorbid SUD / Opposite Action Skill	22 subjects with BPD and SUD in DBT treatment	Non-randomized	10-14 days / User initiated	None, but participants were in DBT outpatient therapy	85% adherence rate with daily assessments. 97% of skill coaching sessions rated helpful. Pre to post improvement in use of opposite action skill ($d = .84$) and confidence with skill ($d = .59$).	Reductions at end of each coaching session on emotional intensity (0-10 scale) ($d = .52$ at end of session, $d = 1.47$ at follow up) and urges to use substances (0-10 scale) ($d = .29$ end of session). Pre to post improvements in psychopathology (BDI) ($d = .55$) and (BSI) ($d = .43$).
<i>Acceptance and Commitment Therapy</i>							
Ly et al., 2014b	Stress in Middle Managers / ACT's Six Basic Tools	Non-clinical sample with ACT app $n = 36$ Waitlist $n = 37$	RCT (ACT app vs. waitlist)	6 weeks / User initiated	Therapist sent encouraging personal text messages every other day	44% adhered to using the app for all 6 weeks.	Greater pre to post improvements on stress (PSS-14 and GHQ-12) with ACT app vs. waitlist. ACT app improved from pre to post on (GHQ-12) ($d = .37$) and (PSS-14) ($d = .62$).

Ly et al., 2012	Feasibility / Values and committed action tools	Non-clinical sample with 11 iPhone users	Non-randomized	4 weeks / User initiated	None	NA	Pre to post improvements on values $d = .77$ (BEVS) and psychological flexibility $d = .50$ (AAQ-II). No significant improvements on psychological symptoms (DASS-21) or life satisfaction (SWL5).
Ahtinen et al., 2013	ACT skills coaching	Non-clinical sample of 15 adults	Non-randomized	4 weeks / 46 User initiated exercises	None	Mean number of usage sessions = 17 (SD = 9, range 5-36). Mean total usage time of 192 minutes (SD = 99, range 56-339).	Pre to post improvements on stress (single item 5 point scale) ($d = 3.8$) and life satisfaction (SWLS) ($d = 2.6$). No improvements on psychological flexibility (AAQ).
Bricker et al., 2014	Smoking cessation / ACT skills and related tools for smoking (e.g., quit plan)	Smoker sample with ACT app $n = 98$ QuitGuide $n = 98$	RCT (ACT app vs. QuitGuide app)	8 weeks / User initiated	None	Used the ACT app more than QuitGuide (37 vs. 15 times). 53% rated ACT app as useful and 59% satisfied with ACT app.	ACT app (quit rate) was non-significantly higher than QuitGuide at two month follow up (13% vs. 8%; OR = 2.7).

*Within condition effect sizes are reported given the high rate of nonrandomized studies and that several RCTs used active treatments in which effect sizes would be more likely to reflect non-inferiority than efficacy. NA = Not applicable due to no available data; cCBT = computer-based CBT; IPT = Interpersonal Psychotherapy; EMI = Ecological Momentary Intervention (prompts received from the app that lead to assessment/tailored intervention); AAQ = Acceptance and Action Questionnaire; SWLS = Satisfaction with Life Scale; DAAS = Depression Anxiety Stress Scale; PSS = Perceived Stress Scale; BDI = Beck Depression Inventory; LSAS = Liebowitz Social Anxiety Scale; GHQ = General Health Questionnaire; BEVS = Bull's Eye Value Survey; BSI = Brief Symptom Inventory; SAD = Social Anxiety Disorder; SUD = Substance User Disorder.

Table 2: Results of searching commercial application stores for cognitive behavioral therapy applications.

Search Term	Apple Store Apps	Google Play Store Apps
“CBT”	206	83
“Cognitive Behavioral Therapy”	21	8
“Cognitive Therapy”	69	27
“Behavioral Activation”	0	0
“DBT”	22	1
“Dialectical Behavioral Therapy”	1	5
“Acceptance and Commitment Therapy”	3	1