Towards managing a platform of mental health apps: A Secondary Analysis of the IntelliCare Field Trial

Mary J. Kwasny\textsuperscript{a}, Stephen M. Schueller\textsuperscript{a1}, Emily Lattie\textsuperscript{a}, Elizabeth Gray\textsuperscript{a}, David C. Mohr\textsuperscript{a}

\textsuperscript{a}Northwestern University, Feinberg School of Medicine, Department of Preventive Medicine. 680 N. Lake Shore Drive, Suite 1400. Chicago, IL 60611 USA.
\textsuperscript{1}Department of Psychology and Social Behavior, University of California, Irvine. Irvine, CA 92697.

m-kwasny@northwestern.edu
schueller@northwestern.edu
emily.lattie@northwestern.edu
elizabeth.gray@northwestern.edu
d-mohr@northwestern.edu

Corresponding author:

Mary Kwasny
Northwestern University
Feinberg School of Medicine
Department of Preventive Medicine
680 N. Lake Shore Drive, Suite 1400
Chicago, IL 60611 USA.
m-kwasny@northwestern.edu
ABSTRACT

**Background:** People using apps for mental health and wellbeing are likely to try multiple apps over time. In general, people use apps to meet immediate needs, and often use a variety of apps to meet larger goals (for example people may have multiple apps to manage various transportation needs). IntelliCare is a mental health app platform that was designed with these principles in mind: the apps are *elemental* in that each app targets a different change strategy; they are *simple and brief* to use; and they are *eclectic*, allowing the user to select which strategies are useful to them. While this may improve engagement, it creates the same challenges faced by users of app stores. Thus, mental health app platforms will require navigation aids, such as recommender systems that can quickly get a person to an app that is useful.

**Objective:** As a first step towards developing navigation and recommender tools, this study explored app use patterns across the IntelliCare platform, and their relationship to depression and anxiety outcomes.

**Method:** This secondary analysis of the IntelliCare Field Trial recruited people with depression and/or anxiety. Participants received 8 weeks of coaching, primarily by text, and weekly recommendations for apps. App use metrics included frequency of use and lifetime use were defined. Depression and anxiety, measured using the PHQ-9 and GAD-7, were assessed at baseline and end of treatment. Ordinal logistic regression models, log-rank tests and cluster analysis were utilized to determine patterns of use, and if these use metrics alone, or in combination, predicted improvement or remission (I/R) in depression or anxiety.

**Results:** The analysis included 96 people with depression and/or anxiety. People generally followed recommendations to download and try new apps each week. Apps clustered into 5 groups: Thinking (apps that targeted or relied on thinking), Calming (relaxation and insomnia), Checklists (apps that used checklists), Activity (behavioral activation and activity), and Other. Both overall frequency of use and lifetime use were predictive in response for depression and anxiety. The Thinking, Calming, and Checklist clusters were associated with improvement in depression and anxiety, and the Activity cluster was associated with improvement in Anxiety only. However, the use of clusters was not more strongly associated with improvement than individual app use.

**Conclusion:** Participants in a field trial remained engaged with a suite of apps for the full eight weeks of the trial. App use patterns did fall into clusters, suggesting knowing something about use of one app may be useful in helping select another app that the person is more likely to use.

**Trial Registration:** Clinicaltrials.gov NCT02176226

**Key Words:** Mobile Applications, Depression, Anxiety
INTRODUCTION
Background

Depression and anxiety are common mental health problems [1,2] that are among the leading causes of morbidity and disability worldwide [3]. The vast majority of those experiencing these common mental health problems cannot access treatment due to a variety of barriers including the lack of availability of services, time constraints, transportation problems, and cost [4,5]. A wide variety of web-based treatments have been developed and shown to be highly effective in the treatment of depression and anxiety, particularly when coupled with some human support to promote adherence and enhance outcomes [6,7]. These programs, leveraging the strengths of computer-accessed web programs in providing information, have strong psychoeducational components along with some interactivity components that function much like worksheets [8].

More recently, mobile apps have been developed and evaluated for the treatment of depression [9,10]. Mobile apps have a number of advantages. Because people keep their phones with them, app-based interventions can fit more seamlessly into the fabric of people’s lives. Smartphones are becoming ubiquitous in developed countries and are increasingly common in developing nations [11].

The design of mobile apps for mental health has posed an interesting challenge compared with in-person and web-based treatments. There are many potential psychological and behavioral strategies that can be used to target specific concerns associated with mental health problems [12,13]. Psychological treatments should be flexible and adaptive, providing the treatment elements that best meet the needs and preferences of the patient [14]. Web-based treatments designed to be delivered via computer can offer a wide variety of treatment approaches, which sometimes require a few layers of navigation. Often these programs require longer periods of engagement with psychoeducational material, with recommended access every week or few days. However, this design may not be well suited to mobile apps. People tend to use apps in very short bursts of time, sometimes frequently [15,16]. Thus, popular apps tend to be designed to be used through simple interactions with limited navigation. A single app tends to focus on a narrow set of objectives. For example, most people do not use one app for transportation needs. Rather, people often use a variety of apps to manage plane flights on different airlines, to manage train transportation and buses, or to map driving routes and check traffic. In short, different apps are used flexibly to meet the changing moment-to-moment needs. Thus, there are a large number of behavioral strategies that may be useful for people with common mental health problems, but mobile apps that are used and useful tend to be narrow in focus and quick to use.

The observation of the incongruity between the large variety of potentially useful psychological strategies and the narrow, efficient design requirements for mobile apps has led our team to question the “app for that” approach that has been common to digital mental health design.
IntelliCare addresses this by creating a suite of apps, each of which addresses a single psychological or behavioral strategy, rather than attempting to address the full theoretical framework for a mental health problem [17]. Thus, each app is elemental, allowing the user to select which strategies are the most useful to them, consistent with the US Institute of Medicine (IOM) report recommending that app-based treatments combine therapeutic elements as well as consider how people tend to use apps. Each app is simple to use, most requiring less than a minute per engagement [18]. Most of the apps focus on supporting learning or implementing a skill, and not on psychoeducation, thus helping people with an immediate problem. The IntelliCare suite of apps, which as of this writing have had more than 85,000 downloads from the Google Play Store, have generally experienced good engagement. A field trial that provided 8 weeks of coaching showed significant reductions in depression and anxiety symptoms.

As we move from a single app that addresses a disorder to a platform of apps that each supports discrete treatment strategies, the management of those apps poses a new challenge. As a mental health app platform such as IntelliCare includes a growing number of apps, user’s reliance on trial-and-error to select apps will make the platform harder to use in a meaningful way. To address this, IntelliCare includes a Hub app, which if downloaded, can help organize the users’ experience by making recommendations about which apps to select. Users who download this Hub app use more apps and use the apps for longer periods of time [17]. Receiving a recommendation from the Hub app to use a specific app increases the likelihood that the app will be downloaded and used [19]. However, it is unclear at this point how to recommend these apps in a way that ensures users are presented with options that are likely to meet their preferences and needs.

Objective

The aim of this study was to investigate the patterns (or clusters) of use across the IntelliCare apps over time to explore whether knowing something about the use of one app might be related to the use of other related apps. We also explored the relationship between app use, as well as the use of app clusters to outcomes of depression and anxiety. This study was aimed not only at understanding the possible utility of apps, but also to examine whether outcomes are driven more by user engagement around a construct (or app cluster), or simply by individual app use. Understanding these use patterns could improve our ability to recommend apps that will be more useful to users.

METHODS

Participants

This is a secondary analysis of a single arm field trial of IntelliCare. Details on the methods are reported in the primary paper [18]. Briefly, participants were recruited from a variety of sources, including online, a healthcare system, community advertising, and research registries. Participants were included in this single arm field trial if they exhibited depressive symptoms indicated by a score of 10 or higher on the Patient Health Questionnaire-9 (PHQ-9) [20], and/or anxiety symptoms indicated by a score of 8 or higher on the Generalized Anxiety Disorder-7 (GAD-7) questionnaire [21]; were 18 years of age or older (age 19 if in Nebraska, given age of
consent); could speak and read English, lived in the United States; and owned and were familiar with an Android smartphone with data and text plans. Participants were excluded if they had any visual, hearing, voice, or motor impairments that would prevent completion of study procedures; reported having a severe psychiatric disorder (e.g., bipolar disorder, psychotic disorder, dissociative disorder) or any other diagnosis for which this trial was either inappropriate or dangerous; exhibited severe suicidality including a plan and intent; had initiated or changed antidepressant and/or anti-anxiolytic pharmacotherapy in the previous 14 days; or had used any of the IntelliCare apps for more than one week in the past three months.

Procedures

People who met inclusion criteria based on online questionnaires and signed an online consent approved by the Northwestern University institutional review board were offered 8 weeks of coaching aimed at helping them use the IntelliCare app suite, detailed below. Outcome assessments, which included the PHQ-9 and GAD-7, were administered at baseline, and weeks 4 and 8.

IntelliCare

At the time of this trial, the IntelliCare platform consisted of 14 apps (Table 1). This included 13 clinical apps, each of which was designed to target a specific behavioral or psychological treatment strategy (e.g. goal setting, behavioral activation, social support, etc.) and improve symptoms of depression and anxiety through efficacious treatment strategies [17]. The user’s experience with the clinical apps was coordinated through a Hub app that, among other functions, made weekly recommendations for new apps. Recommendations were made at random, as there was no basis at the time of this trial to select specific apps. While users were asked to at least to try the newly recommended apps, they were encouraged to use the apps they found most helpful.

Table 1: Usage of all Intelicare Apps

<table>
<thead>
<tr>
<th>Individual App Use</th>
<th>Overall Use</th>
<th>Use by those who downloaded app</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent days used</td>
<td>Lifetime used</td>
</tr>
<tr>
<td></td>
<td>Median (25&lt;sup&gt;th&lt;/sup&gt;, 75&lt;sup&gt;th&lt;/sup&gt;)</td>
<td>Median (25&lt;sup&gt;th&lt;/sup&gt;, 75&lt;sup&gt;th&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Aspire</td>
<td>19 (3, 50)</td>
<td>13 (0, 32)</td>
</tr>
<tr>
<td>Boost Me</td>
<td>12 (0, 25)</td>
<td>7 (0, 22)</td>
</tr>
<tr>
<td>Daily Feats</td>
<td>41 (1, 75)</td>
<td>21 (0, 39)</td>
</tr>
<tr>
<td>iCope</td>
<td>11 (0, 25)</td>
<td>6 (0, 30)</td>
</tr>
<tr>
<td>My Mantra</td>
<td>16 (0, 58)</td>
<td>6 (0, 30)</td>
</tr>
<tr>
<td>ME Locate</td>
<td>2 (0, 17)</td>
<td>0 (0, 8)</td>
</tr>
<tr>
<td>Day to day</td>
<td>26 (4, 57)</td>
<td>16 (0, 36)</td>
</tr>
<tr>
<td>MoveMe</td>
<td>16 (2, 33)</td>
<td>6 (0, 23)</td>
</tr>
</tbody>
</table>
Coaching Protocol

Coaching was guided by the IntelliCare Coaching Manual [22], which is based on aspects of the Efficiency Model of Behavioral Intervention Technologies (BIT) Support [23] and supportive accountability [24]. Coaching was aimed primarily at encouraging participants to try the apps recommended to them through the Hub app. Coaches also answered questions about how to use the tools found in the apps and the rationale behind the skills taught by the apps, encouraged application of the skills in daily life, and provided some technical support as needed. Coaching began with an initial 30-45 minute engagement phone call to establish goals for mood and anxiety management, ensure the participant could download the Hub app from the Google Play store, introduce the suite of available smartphone apps, build rapport, and set expectations for the coach-participant relationship. Some participants also received an additional 10 minute call around mid-treatment. After the initial engagement call, participants received 2-3 text messages per week from their coach to provide support in using apps, offer encouragement, reinforce app use, and check-in on progress or challenges. Coaches also responded to all participant-initiated text messages within 1 working day. The coaches had a dashboard that provided information about the IntelliCare apps on each participant’s phone, including which apps were installed, when they were downloaded, each time an app was used, and which apps were selected as “primary” in the Hub app. The dashboard also included an SMS messaging tool, a section for brief notes, and an alert indicating when no IntelliCare app had been used for 3 days, prompting coaches to check in. Coaches had at least a bachelor’s degree in psychology or a related field and were trained and monitored by one of the coaching manual authors.

Outcome Assessment

Depression was measured using the PHQ-9 [20] and anxiety was measured using the GAD-7 [25], two commonly used self-report measures. Outcomes were categorized using standard cutoffs [26,27]. For depressed patients, we defined remission as PHQ-9 < 5, improvement as 5 ≤ PHQ-9 <10, and no improvement as PHQ-9 ≥ 10. For participants with anxiety, remission was determined as GAD-7 < 5, improvement as 5 ≤ GAD-7 < 8, and no improvement as GAD-7 ≥ 8.
App Use Metrics

Although apps were available for download at any time during the trial, app recommendations were made randomly over the course of 8 weeks, and the time that any individual app was on a participant’s phone could vary substantially. Therefore, in addition to total apps used, other use metrics needed to be defined in relation to the time that the app was on a phone and available to the participant. Accordingly, we created two metrics: frequency of use and lifetime use. Frequency of use was measured as the percentage of days used, calculated as the number of days the app or set of apps was launched divided by the number of days in the study after the app was downloaded by a participant and available for use. For instance, since the trial lasted 8 weeks, or 56 days, if the participant downloaded the app on day 14, the possible number of days it could have been used was 56 minus 14, or 42 days; if they launched the app on 21 of those days, the frequency of use would be 50%. We note that we made an a priori decision to measure app use frequency based on days of use rather than the number of times an app was launched, as we were concerned about the numerous potential sources of variability for app use within a day, such as interruptions, or differences in app design that might encourage differential use. Lifetime use was defined as the time between the first launch and the last launch. However, we recognized that not every app was designed to be used every day, so we allowed for censoring if the app was still being used in the last week of the trial. For instance, if the app was first launched on day 14, and last launched on day 22, the lifetime use of the app would be 8 days. However, if the app was first launched on day 14 and last used in the last week of the study (between days 49 and 56), the lifetime use would be censored at 56 minus 14, or 42 days.

Statistical Analyses

We describe app use using visual informatics and descriptive statistics. Additionally, we performed a cluster-analysis on the total number of launches by app using a centroid approach with Spearman’s correlations to determine if the use of any of the apps was clustered together.

Analyses relating app use to outcome were conducted separately for participants who met the PHQ-9 ≥10 entry criterion for depression and participants who met the GAD-7 ≥ 8 criterion for the anxiety. We fit ordinal logistic regression models, modeling proportional odds of improvement or remission (I/R) to determine if frequency of use, for any apps combined, individual apps, and clusters of apps varied by outcome. To assess if use of multiple apps within a cluster was more effective in treating depression or anxiety, we “scored” apps if they had 25% frequency of use or more, and then examined if there was a trend between the number of scored apps within a cluster and outcome. Lastly, we employed stepwise selection to determine which apps or combination of apps were most predictive in a model on our ordinal outcomes concerning the I/R of symptoms. All odds ratios (OR) and 95% confidence intervals (CI) are presented for a 10% increase in days used. To examine lifetime use by outcome, we fit Kaplan-Meier plots and used log-rank tests.

All analyses were run in SASv9.4 (Cary, NC); graphs were created in SASv9.4 or Rv3.4.3 [28]. The type I error was set at .05 for all analyses, but we caution that any findings should be further investigated, as this is a secondary analysis and therefore subject to increased type I errors.
RESULTS

Participants

A detailed description of the participants and primary outcomes have been published elsewhere [18]. Briefly, 99 participants were enrolled and began the 8-week field trial. A flow diagram is available in the main outcome paper [18], however only 3 participants were lost to follow-up, leaving 96 participants with at least two outcome assessments, who are the focus of this secondary analysis. At the start of the field trial, the depression criterion of PHQ-9 ≥10 was met by 78 (81%) participants, the anxiety criterion of GAD-7 ≥8 was met by 77 (80%) participants, and 59 (61%) met criteria for both depression and anxiety. Using our classification of treatment response, among the 78 participants meeting the entry criterion depression, by the end of treatment 26 (33%) were in remission, 30 (38%) had improved, and 22 (28%) remained symptomatic. Of the 77 participants who met criteria for anxiety, 28 (36%) were in remission, 23 (30%) improved, and 26 (34%) remained symptomatic.

General App Use

Suite Use

Participants downloaded an average of 9.3 apps (SD=2.6) over the 8 weeks; half of the participants used over 9 apps. Overall, participants tended to download the apps in the suite gradually over the entire 8 weeks, rather than download all of the apps at the start of the trial. There was an immediate launch of a few apps in the first week (average number downloaded 2.4 (SD 2.0)), which tapered gradually over the length of the study with average increases of 1.4 apps/week for weeks 2 and 3, and roughly 1 app/week for weeks 4-6, down to 0.5 apps/week for weeks 7 and 8 (Figure 1). The apps were used throughout the trial, with the median lifetime use for the suite of 55 days (Interquartile range 53-55).
Relationship of suite use to Outcome

The total number of apps used was not associated with I/R in depression (OR=1.07 (95% CI 0.91, 1.25)), but was associated with I/R in anxiety (OR=1.20 (95% CI 1.01, 1.43)). Conversely, the percentage of days with any app use was associated with the odds of I/R of depression (OR=1.26 (95% CI 1.05, 1.52)), while it was not associated with I/R of anxiety (OR 1.19 (0.99, 1.43)) (Table 2).

Table 2: Ordinal Odds Ratios (OR) for Improvement or remission (IR) of Depression/Anxiety by App Use

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
<th>Anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariate OR for Suite</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Apps downloaded</td>
<td>1.07 (0.91, 1.25)</td>
<td>1.20 (1.01, 1.43)</td>
</tr>
<tr>
<td>Percent days of app use</td>
<td>1.26 (1.05, 1.52)</td>
<td>1.19 (0.99, 1.43)</td>
</tr>
<tr>
<td><strong>Univariate OR for Individual Apps</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspire</td>
<td>1.15 (1.00, 1.32)</td>
<td>1.19 (0.99, 1.43)</td>
</tr>
<tr>
<td>Boost Me</td>
<td>1.27 (1.00, 1.61)</td>
<td>1.37 (1.06, 1.77)</td>
</tr>
</tbody>
</table>
Individual App Use

App Frequency

Half of the participants used at least one IntelliCare treatment app for 49 (88%) of the 56 days, with interquartile range of 35 (63%)-53 (95%) days. Median percent of days used for specific apps ranged from 2% (Me Locate) to 41% (Daily feats) (Table 1).

Relationship of Frequency to Outcome

For depression outcomes, use of Aspire was associated with increased odds of I/R (OR 1.15 (1.00, 1.32)), and the use of Purple Chill was associated with increased odds of I/R, but the assumption of proportional odds was not met, and increased odds of remission was seen only among those who showed improvement (OR 1.39, (1.11, 1.72). For anxiety outcomes, use of BoostMe (1.37 (1.06, 1.77)), Day-to-Day (1.82 (103, 1.35)), Me Locate (1.82 (1.25, 2.63)), Purple Chill (1.16 (1.01, 1.34)), and Thought Challenger (1.15 (1.01, 1.31)) were all associated with increased odds of I/R (Table 2).

| Daily Feats   | 1.09 (0.97, 1.22) | 1.05 (0.94, 1.18) |
| Day to day    | 1.11 (0.98, 1.27) | 1.18 (1.03, 1.35) |
| iCope         | 1.00 (0.86, 1.16) | 1.01 (0.86, 1.17) |
| My Mantra     | 1.02 (0.91, 1.15) | 1.13 (0.99, 1.28) |
| ME Locate     | 1.15 (0.93, 1.41) | 1.82 (1.25, 2.63) |
| MoveMe        | 1.15 (0.99, 1.34) | 1.17 (0.99, 1.37) |
| Purple Chill  | 1.17 (1.02, 1.34) | 1.16 (1.01, 1.34) |
| Slumber time  | 1.04 (0.92, 1.18) | 1.07 (0.94, 1.23) |
| Social Force  | 1.05 (0.89, 1.25) | 1.16 (0.96, 1.42) |
| Thought Challenger | 1.07 (0.93, 1.22) | 1.15 (1.01, 1.31) |
| Worry Knot    | 1.04 (0.90, 1.21) | 1.06 (0.91, 1.23) |

Univariate OR for Clusters

| Thinking      | 1.16 (1.01, 1.33) | 1.17 (1.02, 1.35) |
| Calming       | 1.17 (1.02, 1.35) | 1.18 (1.02, 1.36) |
| Checklist     | 1.18 (1.04, 1.34) | 1.14 (1.00, 1.30) |
| Activity      | 1.08 (0.92, 1.27) | 1.26 (1.05, 1.52) |
| Other         | 1.17 (0.99, 1.37) | 1.16 (0.99, 1.36) |

Multivariate Models

| BoostMe       | 1.31 (1.02, 1.67) |
| Lifetime Relax | 1.25 (1.07, 1.47) |
| MeLocate      | -- (1.74 (1.19, 2.55) |
| Lifetime Thought Challenger | -- (1.20 (1.02, 1.43) |

aAs Lifetime is measured in days, and one extra day of use is clinically meaningless, Odds Ratios are based on a unit difference of one week.
Lifetime Use

The lifetime use for specific apps ranged from 0 days (MeLocate) to 21 days (Daily feats). Kaplan-Meier plots for each app’s lifetime use over the trial period are shown (Figure 2).

![Kaplan-Meier plots for each app's lifetime use over the trial period](image)

Figure 2: Lifetime Use of each App

**Relationship of Lifetime use to Outcome**

There were significant differences in the lifetime use for iCope, MoveMe, and PurpleChill, between participants who had remission, improvement, or showed no improvement for depression (Log-rank $P=.02, .04, \text{ and } .02$, respectively) (Figure 3a). Among those with anxiety, lifetime use of Boost Me, My Mantra, Me Locate, MoveMe, ThoughtChallenger, and WorryKnot varied across remission, improvement, no improvement status ($Ps=.001, .02, .001, .04, .008, \text{ and } .04$, respectively) (Figure 3b).
Clusters of Apps

A cluster analysis revealed 5 groups of apps that tended to be used together. The average within cluster R-squared ranged from .34 to .61, whereas the average R-squared to the nearest cluster was .03 to .33. In total the 5 clusters explained 59% of the variability of correlations between launches. The five clusters identified could best be described as:

“Thinking” - Thought Challenger, MyMantra, Day to Day, and iCope;
“Calming” - Purple Chill and Slumbertime;
“Checklists” - Aspire and Daily Feats;
“Activity” - BoostMe and MoveMe; and
“Other” - MeLocate, Social Force, and Worry Knot, which appeared unified only in the lack of engagement relative to other apps.

Use

The Thinking and Checklist clusters were used the most often, and for a longer period of time than the other clusters, with median lifetimes over 5 weeks in the 8 week trial. The Calming and Activity clusters were also used fairly often with a median use of just over 4 weeks. The Other cluster was used the least, with a median lifetime of a week (Table 1). The number of apps
scored (≥ 25% frequency of use) in each cluster was fairly diverse for the Thinking, Calming, and Checklist clusters, where there were participants who did not use any of the apps very much as well as participants who used all of the apps within the cluster. The Activity and the Other clusters were more prone to have individuals who did not use any of the apps very often (Table 3).

Table 3: Number of apps with more than 25% frequency of use within each App Cluster

<table>
<thead>
<tr>
<th>Number of apps with 25% use</th>
<th>Thinking</th>
<th>Calming</th>
<th>Checklist</th>
<th>Activity</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25 (26%)</td>
<td>29 (30%)</td>
<td>23 (24%)</td>
<td>52 (54%)</td>
<td>49 (51%)</td>
</tr>
<tr>
<td>1</td>
<td>21 (22%)</td>
<td>36 (38%)</td>
<td>40 (42%)</td>
<td>32 (33%)</td>
<td>31 (32%)</td>
</tr>
<tr>
<td>2</td>
<td>22 (23%)</td>
<td>31 (32%)</td>
<td>33 (34%)</td>
<td>12 (13%)</td>
<td>12 (13%)</td>
</tr>
<tr>
<td>3</td>
<td>17 (18%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4 (4%)</td>
</tr>
<tr>
<td>4</td>
<td>11 (11%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Relationship of Cluster use to Outcome

Among participants entering the study with high levels of depressive symptoms, the odds of I/R were significantly greater for participants with a higher percentage of app use days for apps in the Thinking cluster (OR=1.16, 95%CI=1.01-1.33), the Calming cluster (OR=1.17, 95%CI = 1.02-1.35), and the Checklist cluster (OR=1.18, 95% CI 1.04-1.34) (Table 2). We also examined whether the number of apps within a cluster was related to outcome. There was a significant trend in I/R when multiple apps were used in the Checklist cluster (OR=1.96, 95% CI 1.12-3.44), as well as in the Activity cluster (OR=1.81, 95% CI 1.00-3.26). However, in stepwise multivariable ordinal regression models, after adjusting for frequency of use in the Checklist cluster, no other cluster use nor number of apps used within cluster increased the odds of I/R.

Among participants entering with high levels of anxiety symptoms, the odds of I/R were significantly greater for participants with a higher frequency of use in the Thinking cluster (OR=1.17, 95%CI=1.02-1.35), the Calming cluster (OR=1.18, 95% CI 1.02-1.36), the Checklist cluster (OR= 1.14, 95%CI = 1.00-1.30); and the Activity cluster (OR=1.26, 95% CI 1.05-1.52) (Table 2). Additionally, there was a significant trend in I/R as the number of apps used increased within the Thinking cluster (1.47 (1.07-2.02)), the Checklist cluster (1.79 (1.03-3.13)), the Activity cluster (1.87 (1.03-3.39)), and the Other cluster (OR 2.19 (1.24-3.89)). After adjusting for frequency of use in the Activity cluster, no other cluster nor number of apps used within cluster increased the odds of I/R in multivariable ordinal regression models.

For depression, longer lifetime use of the Activity cluster was significantly associated with better outcomes (Log Rank P=0.02). For anxiety, longer use of the Calming, Activity, and Other clusters was associated with better outcomes (Log Rank Ps = .01, .03, and .002, respectively) (Figure 4).
Optimal Use patterns

For participants with depression, the percentage of days using BoostMe in combination with lifetime use of Purple Chill was the most predictive combination of use for response. After adjusting for percent of days using BoostMe, and the lifetime use of Purple Chill, no other app or cluster use metric was significantly associated with I/R. For participants with anxiety, the use of MeLocate and lifetime use of Thought Challenger was the most predictive use combination of apps for I/R. After adjusting for these, no other app or cluster use metric was, significantly associated with I/R (Table 2).
DISCUSSION

Principal Results

Determining the definition of “use” for apps can be challenging. Here we defined two metrics: frequency (percent of days used) and lifetime use (time between initial and last launch). Additionally, we examined clusters of app use based on correlations between the total number of launches, and were able to identify groups of apps that could be defined based on behavioral strategy and user interaction style. This revealed five clusters: the Thinking cluster included apps that prompt or rely on a person to use cognitive processes. The Calming cluster provided tools for relaxation and strategies to improve sleep. The Checklist cluster was defined by the type of interaction people had with the app – the use of checklists – rather than by a psychological strategy. This underscores that the design and interaction features used in apps may be as important to people’s preferences as the psychological goal or behavioral strategies. The Activity cluster was defined by apps that targeted behavioral activation and physical activity. A fifth cluster, which we called Other, consisted of apps that may need further development. Two out of the three had the lowest use, and the third, Worry Knot, had an interaction design that was often not well received, based on user feedback. Nonetheless, the fact that the clusters based on use were well defined suggests that recommendation systems could be useful in getting apps to people that they are more likely to use.

We explored the relationship individual apps and outcome. For depression, Purple Chill (relaxation) and Aspire (living one’s values) were predictive of improvement, while Boost Me (behavioral activation), Day to Day (psychoeducation), Me Locate (used geofencing), Purple Chill, and Thought Challenger (cognitive restructuring) were associated with improvement in anxiety. Given app use was in the context of a suite, it is difficult to interpret these findings. It is likely that improvement is not necessarily due to the use of “an app,” rather a mix of apps.

One hypothesis we had was targeting a construct (e.g. Thinking or Calming) through use of a set of apps may be more beneficial than the use of any individual app. Indeed, the Thinking, Calming, and Checklist clusters were all related to improvement in depression and those three clusters along with Activity were associated with improvement in anxiety. However, our stepwise models did not conform to this hypothesis. When individual apps and clusters were analyzed together, the frequency of Boost Me and lifetime use of Purple Chill were associated with improvement in depression, while the frequency of MeLocate and lifetime use of Thought Challenger were associated with improvement in anxiety. Thus, our hypothesis regarding the use of clusters for improvement received partial support, but was not robust enough in the presence of all app use data.

Finally, we note that engagement remained high throughout the study. We have noted previously that in the public deployment of IntelliCare through the Google Play Store, providing recommendations increases the likelihood that an individual will download an app [19]. However, it was unclear how people would use apps in the context of a treatment where all apps where available from the start. We see here that people tended not to download all of the apps at once, but rather they waited for the weekly recommendations to download and initiate use.
This study provides the first view of how digital mental health platforms that provide a variety of apps or treatments may be optimized. These findings suggest that knowing something about a person’s use of one app may be helpful in selecting the next app to recommend. There is some support for the idea that use of clusters may also be helpful in improving symptoms, although these findings were not robust. Together, these findings support the idea that recommendation engines may be useful in promoting use in platforms with multiple apps such as IntelliCare, and may also be helpful in promoting symptom improvement. This will be critical since, to maintain engagement with an app platform, it will be important to connect people to apps quickly that they want to use.

Limitations

There are limitations in this study that should be considered in interpreting these findings. Chief among them, we have performed a large number of analyses for the sample size. Thus, some of findings may be spurious, and in other cases we were likely underpowered for the number of variables included in analyses.

Conclusion

We saw that a suite of apps was engaging to participants in a field trial for treatment of depression or anxiety. Despite all apps being available for immediate download, participants gradually downloaded and engaged with various apps throughout the trial.

Although the use metrics of different apps in the suite are correlated, a stepwise analysis showed that the use of BoostMe (an Activity focused app) and sustained use of Purple Chill (a Calming app) were the most effective at improving depression, while the use of MeLocate and sustained use of Thought Challenger (a Thinking app) were the most effective at improving anxiety.

Acknowledgements

Research reported in this publication was supported by the United States National Institute of Mental Health grants R01 MH100482 to Dr Mohr, R01 MH109496 to Drs Cheung and Mohr, K08 MH102336 to Dr Schueller, and the National Center for Advancing Translational Sciences, Grant Number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

Dr. Mohr has accepted consulting fees from Optum Behavioral Health, and has an ownership interest in Actualize Therapy. Dr. Lattie has accepted consulting fees from Actualize Therapy. None of the other authors have any conflicts of interest to declare.

References


