JAMA Network Open.

Original Investigation | Psychiatry

A Machine Learning Approach to Understanding Patterns of Engagement With Internet-Delivered Mental Health Interventions

Isabel Chien, MEng; Angel Enrique, PhD; Jorge Palacios, MD, PhD; Tim Regan, PhD; Dessie Keegan, MSc; David Carter, PhD; Sebastian Tschiatschek, PhD; Aditya Nori, PhD; Anja Thieme, PhD; Derek Richards, PhD; Gavin Doherty, DPhil; Danielle Belgrave, PhD

Abstract

IMPORTANCE The mechanisms by which engagement with internet-delivered psychological interventions are associated with depression and anxiety symptoms are unclear.

OBJECTIVE To identify behavior types based on how people engage with an internet-based cognitive behavioral therapy (iCBT) intervention for symptoms of depression and anxiety.

DESIGN, SETTING, AND PARTICIPANTS Deidentified data on 54 604 adult patients assigned to the Space From Depression and Anxiety treatment program from January 31, 2015, to March 31, 2019, were obtained for probabilistic latent variable modeling using machine learning techniques to infer distinct patient subtypes, based on longitudinal heterogeneity of engagement patterns with iCBT.

INTERVENTIONS A clinician-supported iCBT-based program that follows clinical guidelines for treating depression and anxiety, delivered on a web 2.0 platform.

MAIN OUTCOMES AND MEASURES Log data from user interactions with the iCBT program to inform engagement patterns over time. Clinical outcomes included symptoms of depression (Patient Health Questionnaire-9 [PHQ-9]) and anxiety (Generalized Anxiety Disorder-7 [GAD-7]); PHQ-9 cut point greater than or equal to 10 and GAD-7 scores greater than or equal to 8 were used to define depression and anxiety.

RESULTS Patients spent a mean (SD) of 111.33 (118.92) minutes on the platform and completed 230.60 (241.21) tools. At baseline, mean PHQ-9 score was 12.96 (5.81) and GAD-7 score was 11.85 (5.14). Five subtypes of engagement were identified based on patient interaction with different program sections over 14 weeks: class 1 (low engagers, 19 930 [36.5%]), class 2 (late engagers, 11 674 [21.4%]), class 3 (high engagers with rapid disengagement, 13 936 [25.5%]), class 4 (high engagers with moderate decrease, 3258 [6.0%]), and class 5 (highest engagers, 5799 [10.6%]). Estimated mean decrease (SE) in PHQ-9 score was 6.65 (0.14) for class 3, 5.88 (0.14) for class 5, and 5.39 (0.14) for class 4; class 2 had the lowest rate of decrease at −4.41 (0.13). Compared with PHQ-9 score decrease in class 1, the Cohen *d* effect size (SE) was −0.46 (0.014) for class 2, −0.46 (0.014) for class 3, −0.61 (0.021) for class 4, and −0.73 (0.018) for class 5. Similar patterns were found across groups for GAD-7.

CONCLUSIONS AND RELEVANCE The findings of this study may facilitate tailoring interventions according to specific subtypes of engagement for individuals with depression and anxiety. Informing clinical decision needs of supporters may be a route to successful adoption of machine learning insights, thus improving clinical outcomes overall.

JAMA Network Open. 2020;3(7):e2010791. doi:10.1001/jamanetworkopen.2020.10791

Key Points

Question Can machine learning techniques be used to identify heterogeneity in patient engagement with internet-based cognitive behavioral therapy for symptoms of depression and anxiety?

Findings In this cohort study using data from 54 604 individuals, 5 heterogeneous subtypes were identified based on patient engagement with the online intervention. These subtypes were associated with different patterns of patient behavior and different levels of improvement in symptoms of depression and anxiety.

Meaning The findings of this study suggest that patterns of patient behavior may elucidate different modalities of engagement, which can help to conduct better triage for patients to provide personalized therapeutic activities, helping to improve outcomes and reduce the overall burden of mental health disorders.

➕ **Invited Commentary**

➕ **Supplemental content**

Author affiliations and article information are listed at the end of this article.

Introduction

The World Health Organization defines health as a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.[1] Mental disorders present a substantial burden for good health as they have deleterious effects on the individual, society, and the worldwide economy,[2-4] making their prevention and treatment a public health priority.[5-7]

Responding to the demand for accessible and sustainable mental health care services, internet-delivered psychological interventions offer access to evidence-based treatment and positive clinical outcomes while maintaining quality of care and reducing costs.[8,9] Extensive research has reported possible effectiveness of these interventions for treating psychological disorders.[9-13] However, more complete understanding of the clinical use of digital therapy programs requires further research.[14-16] Most previous studies explored the association between use of the interventions and outcomes, relying on single metrics, such as raw use counts.[17,18] Other studies suggest that single metrics are unlikely to sufficiently capture associations between engagement and outcomes, especially when compared with other factors, such as the actual level of attention or interactivity during an intervention.[19,20] Thus, identifying different behavioral patterns of engagement and linking these patterns to clinical outcomes offer new opportunities for personalizing treatment delivery to reduce nonadherence to therapy and enhance possible effectiveness.[20,21]

The aim of this study was to examine whether different types of patient behaviors manifest in the way people engage with an internet-based cognitive behavioral therapy (iCBT) intervention for symptoms of depression and anxiety. We used machine learning to build a probabilistic graphical modeling framework to understand longitudinal patterns of engagement with iCBT.[22-24] We hypothesized that these patterns would allow us to infer distinct, heterogeneous patient behavior subtypes. We further hypothesized that these subtypes are associated with the intervention's success of improving mental health and that different subtypes of engagement are associated with differences in clinical outcomes.

Methods

Study Design

We used clinical measures and behavioral engagement data from SilverCloud Health. SilverCloud Health is an evidence-based, online, self-administered platform that delivers iCBT alongside feedback from trained human supporters.[25,26] We used deidentified data from 67 468 patients on the Space From Depression and Anxiety treatment program between January 31, 2015, and March 31, 2019. We removed 12 864 individuals who had no supporter assigned and restricted analysis to the remaining 54 604 patients who viewed the program content at least once. The program consists of 8 core modules covering the CBT principles for treating symptoms of depression and anxiety. Content is delivered using textual and audiovisual materials, interactive tools, and personal stories. The platform includes several interactive tools, such as journal, quizzes, mood trackers, and other CBT-based exercises. Human supporters provide guidance to patients in the first 8 weeks of treatment. Further details of the platform and tools are available in the eMethods and eTable 1 in the Supplement. Data analysis was carried out between April 1 and October 31, 2019. All users provided written or oral consent for their anonymized data to be used in routine evaluations for service monitoring and improvement. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cohort studies. Per the Common Rule, institutional review board review was not required for this study, which used deidentified publicly available data.

Clinical Outcomes

We assessed symptoms of depression using the Patient Health Questionnaire-9 (PHQ-9) and symptoms of anxiety using the Generalized Anxiety Disorder-7 (GAD-7). We used a PHQ-9 cut point

of greater than or equal to 10^{27} and GAD-7 scores greater than or equal to 8 to define depression and anxiety cases.[28] These measures were collected at baseline and during routine outcome monitoring (biweekly) up to 14 weeks. A decrease of greater than or equal to 6 on PHQ-9 and greater than or equal to 4 points on GAD-7 represent clinical improvement. The rate of reliable improvement was calculated based on the Reliable Change Index (RCI) using Jacobson and Truax criteria.[29]

We defined 2 types of engagement: whether a patient used the program in a given week (yes/no) and whether a patient used a particular section of the program in a given week. There are 14 sections in total as described in eTable 2 in the Supplement.

Comparing these 2 modalities of engagement allowed us to meaningfully operationalize a measure of engagement: Is it more important that patients engage with the program, or does engagement with specific treatment sections also matter? Distinguishing the iCBT components that patients engage with may allow us to better motivate engagement with therapy based on increased understanding of their behavior.

We developed a probabilistic latent variable model to infer distinct subtypes of patients based on their interaction with the depression and anxiety program over 14 weeks, which was the median time spent with the treatment.

We assumed n latent classes, which represent these subtypes/patterns of engagement. The size and number of these latent classes was unknown a priori, but we learned the optimal number of latent classes (between 2 and 10) by comparing model loss functions. We explored 2 probabilistic model formulations to infer the number of latent subtypes that best encapsulate longitudinal heterogeneity of patterns of engagement: a mixture of the hidden Markov model (**Figure 1**; eMethods in the Supplement) and a latent variable mixed model. For the hidden Markov model, we assumed that weekly observed engagement is encoded by a latent state while the conditional dependence structure on the longitudinal sequence of states is governed by an overarching discrete latent state that represents prototypical patterns of engagement over time (Figure 1). The latent variable mixed model assumes that behavior within a latent class depends on a time parameter rather than on dynamic transitions between states.

Figure 1. The Hidden Markov Model
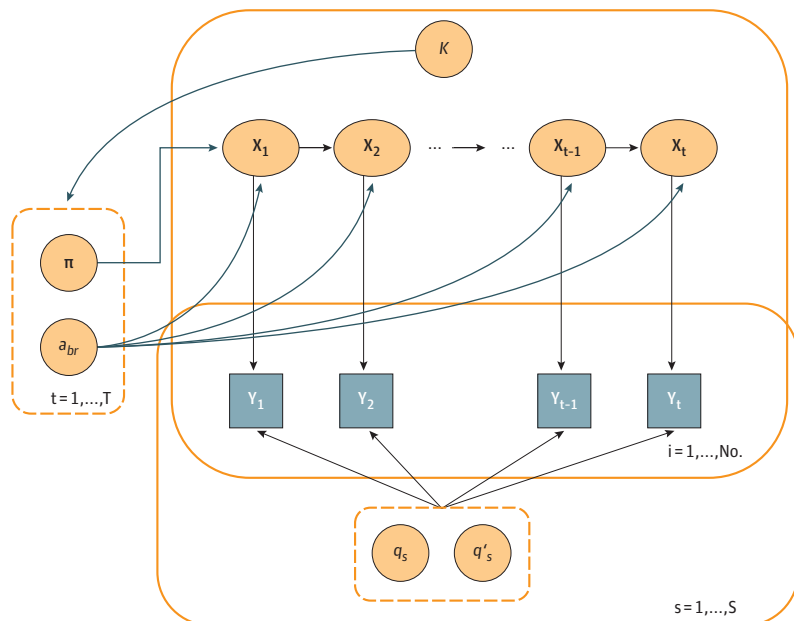
Y represents a binary response (0 = no, 1 = yes) to the question, "Has patient i engaged with the platform at time t?" where t takes values from 1 to 14. K is a multinomial latent variable per patient and is set in ranges of 2 to 10, π is the starting probability, a_{br} is the class (K)–dependent transition probabilities of transition from state b to state r, q_s and q'_s are the emission probabilities of observing engagement with a section () conditioned on the hidden state, q_s indicates that the hidden state takes a value of 1, and q'_s indicates that the hidden state takes a value of 0. When sections data are used, emission probabilities are considered distinct for each section.

Statistical Analysis

We assessed whether there were specific iCBT sections to which different subgroups had a particular affinity and whether subgroups were associated with different profiles of PHQ-9 and GAD-7 using longitudinal linear regression techniques. Tool use was evaluated within the first 2 weeks on the program to investigate whether we could identify early predictors of subsequent engagement patterns of how people used the program. These patterns of tool use were not used to identify the engagement classes themselves, but served as an external validation to assess whether different classes based on program engagement alone could also indicate what tools people in particular subgroups were more likely to interact with. Analysis was carried out using Python, version 3.7 (Python Software Foundation), and Stata, version 15.1 (StataCorp LLC). Findings were considered significant at $P \le .05$.

Results

We used data from a total of 54 604 patients. Patients spent a mean (SD) of 111.33 (118.92) minutes on the program, used 230.60 (241.21) tools, and had baseline scores of 12.96 (5.81) in PHQ-9 and 11.85 (5.14) in GAD-7. Across all patients, clinical scores improved over 14 weeks by a mean (SD) of 4.29 (4.90) for PHQ-9 and 4.01 (4.61) for GAD-7.

Figure 2 shows the distribution of engagement over time based on results from a hidden Markov model, which gave the best model fit. The model identified 5 subtypes of engagement (latent classes) based on patient interaction with sections over time rather than engagement as a binary measure at each week. **Figure 3** illustrates characteristics of different subtypes in terms of the number of modules completed, mean time spent, and number of sessions per week. **Table 1** reports the odds ratio (OR) of each class compared with all other classes using a particular tool conditioned on all other tools used in the first 2 weeks.

Class 1 (low engagers: 19 930 [36.5%]) had the lowest probability of engagement with a steady dropout over time. They spent less time on the program, used fewer sessions, and completed fewer modules compared with all other classes. This class interacted more with the review page where users can communicate with their supporter. With regard to early use, when compared with all other groups, low engagers were more likely to use sections associated with mood monitoring and worry (Worry Tree: OR, 1.26; 95% CI, 1.04-1.52; $P = .02$; My Worries: OR, 1.34; 95% CI, 1.12-1.61; $P = .002$; Anxious Thoughts and Worry Quiz: OR, 1.45; 95% CI, 1.21-1.74; $P < .001$; Mood Monitor: OR, 1.23; 95% CI, 1.05-1.45; $P = .01$; and Understanding My Situation: OR, 1.22; 95% CI, 1.04-1.42; $P = .001$), but less likely to do activities such as Activity Scheduling (OR, 0.57; 95% CI, 0.48-0.68; $P < .001$) and Activities List (OR, 0.58; 95% CI, 0.49-0.69; $P < .001$).

Class 2 (late engagers: 11 674 [21.4%]) had initially low engagement with a slower rate of disengagement over time. During the first 2 weeks, they were less likely to use the To-do list at the end of modules and less likely to look at content such as Hierarchy of Fears (OR, 0.56; 95% CI, 0.34-0.91; $P = .02$), Worry Tree (OR, 0.80; 95% CI, 0.66-0.98; $P = .03$), and Graded Exposure Quiz (OR, 0.58; 95% CI, 0.38-0.93; $P = .02$). However, late engagers were more likely to engage with sections associated with Anxiety: Myths and Facts (OR, 1.67; 95% CI, 1.30-2.16; $P < .001$), Staying in the Present (OR, 1.17; 95% CI, 1.00-1.36; $P = .04$), and Stress Response (OR, 1.40; 95% CI, 1.00-1.96; $P = .04$).

Class 3 (high engagers with rapid disengagement: 13 936 [25.5%]) had the sharpest rate of disengagement despite initial high engagement. Patients were more likely to engage with Take Home Points. They were associated with early higher use of tools included in the core modules Understanding Feelings, Spotting Thoughts, Challenging Thoughts, and Boosting Behaviour (eg, Thoughts, Feelings, and Behaviour cycles: OR, 1.31; 95% CI, 1.16-1.47; $P < .001$). However, high engagers with rapid disengagement were also the least likely to engage with the Activity Goals (OR, 0.39; 95% CI, 0.24-0.64; $P < .001$).

Class 4 (high engagers with moderate decrease: 3258 [6.0%]) had a constantly high probability of engaging with the program. Along with class 5, class 4 undertook significantly more sessions, spent more time on the program, and did more modules and tools per week compared with all other classes. They also engaged more with Progress Points, Profile, and Review sections. They were more likely to interact with goal-based tools and quizzes that require more introspection (What's Your Lens? What's Your Thinking Style?: OR, 1.28; 95% CI, 1.02-1.60; *P* = .03; Core Beliefs Quiz: OR, 1.93;

Figure 2. Five Distinct Patterns of Engagement With Internet-Based Cognitive Behavioral Therapy Identified
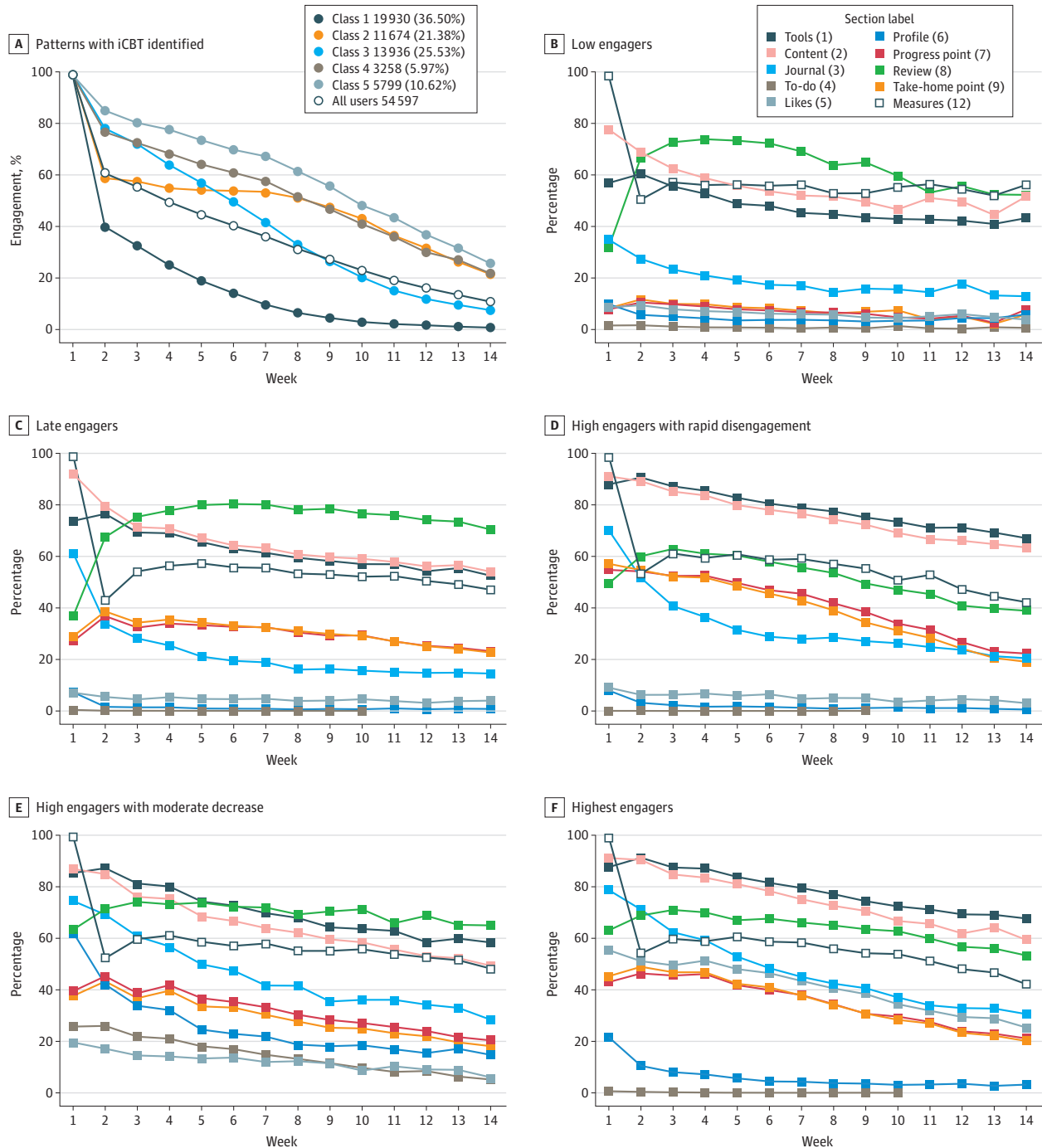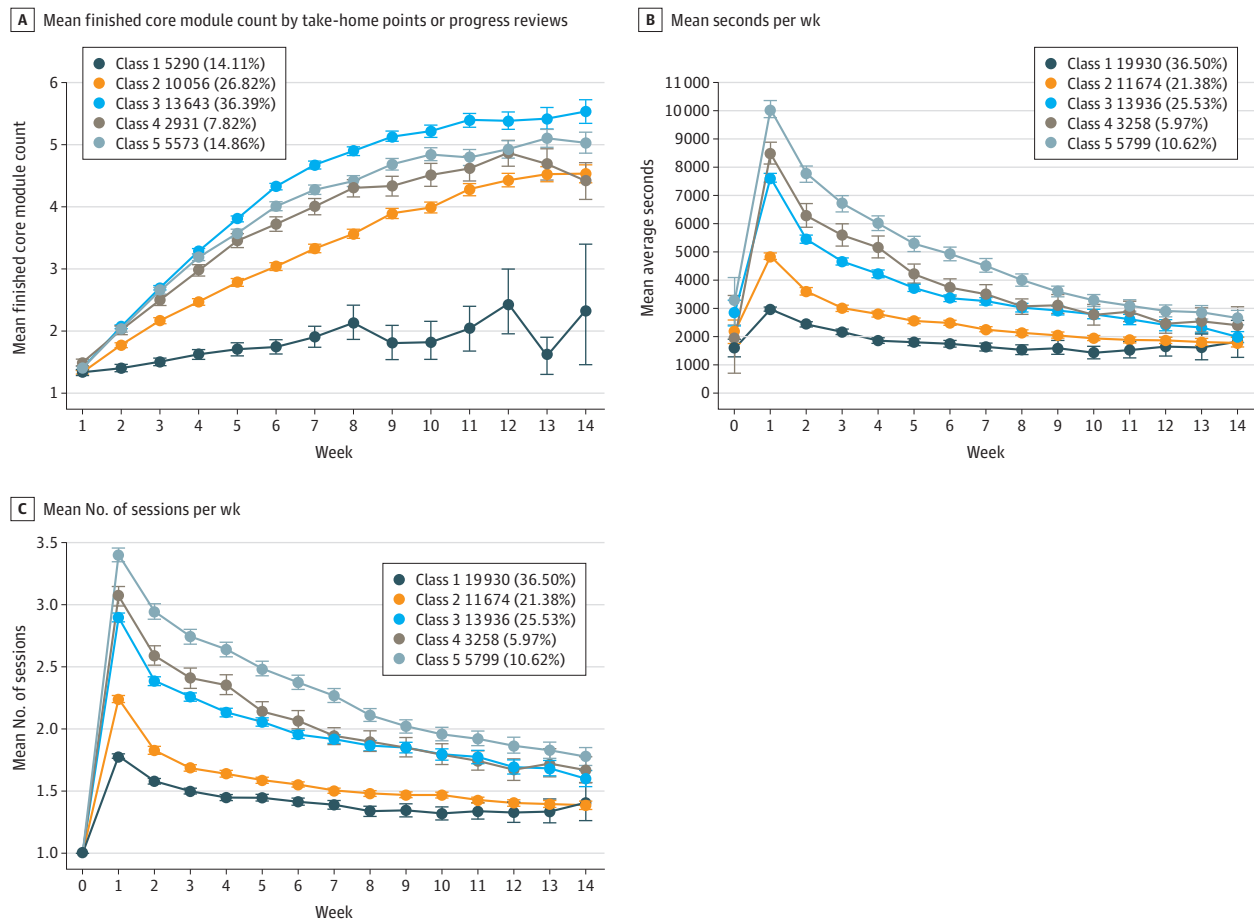

Panel A shows the proportion of users who are engaged for each class at each time point over a 14-week time period. Panels B-F show the proportion of users who use each of the program sections.

95% CI, 1.19-3.16; *P* = .008; and Graded Exposure Quiz: OR, 1.62; 95% CI, 1.06-2.48; *P* = .03), but less likely to use sections containing more reading material (Depression: Myths and Facts: OR, 0.44; 95% CI, 0.27-0.74; *P* = .002; and Anxiety: Myths and Facts: OR, 0.40; 95% CI, 0.24-0.66; *P* < .001).

Class 5 (highest engagers: 5799 [10.6%]) had the highest probability of engagement throughout the time spent on the program. They had significantly more interaction with the journal and were more likely to engage with the sections Progress Points and Take Home Points. Early use indicated ORs above 1 for most modules. The distinguishing feature of this subgroup is that they used Anxious Thoughts and Worries less (OR, 0.82; 95% CI, 0.68-0.99; *P* = .04), but interacted more with Sleeping Tips (OR, 1.47; 95% CI, 1.16-1.86; *P* = .002) and Relaxation (OR, 2.12; 95% CI, 1.57-2.86; *P* < .001), which are sections that require unlocking by a human supporter.

The reliable improvement for class 1 was 39.5%; class 2, 54.8%; class 3, 58.0%; class 4, 58.8%; and class 5, 66.9%. We assumed that PHQ-9 scores were missing at random conditional on individual engagement class membership. To test this missingness assumption and ensure robustness, we restricted analysis to users who had completed 3 or more PHQ-9 assessments (n = 31 466). We found consistent estimates in terms of the estimated longitudinal mean improvement in symptoms of depression and anxiety within each subgroup (eTable 3 in the Supplement). **Table 2** summarizes the estimated class-specific change in PHQ-9 and GAD-7 scores over a 14-week period. Patients in class 3 spent less time on the program than those in classes 4 and 5 but had significantly greater weekly change in PHQ-9 scores, with an estimated mean (SE) decrease in PHQ-9 score of 6.65 (0.14)

Figure 3. Patterns of Interaction With the Program for Engagement Subtypes

A Mean finished core module count by take-home points or progress reviews

Class 1 5290 (14.11%)
Class 2 10056 (26.82%)
Class 3 13643 (36.39%)
Class 4 2931 (7.82%)
Class 5 5573 (14.86%)

B Mean seconds per wk

Class 1 19930 (36.50%)
Class 2 11674 (21.38%)
Class 3 13936 (25.53%)
Class 4 3258 (5.97%)
Class 5 5799 (10.62%)

C Mean No. of sessions per wk

Class 1 19930 (36.50%)
Class 2 11674 (21.38%)
Class 3 13936 (25.53%)
Class 4 3258 (5.97%)
Class 5 5799 (10.62%)

Interactions based on core modules completed (A), mean time spent per week on the program (B), and mean number of sessions per week (sum of session duration) (C). Error bars indicate 95% CIs.

Table 1. Tools Used in the First 2 Weeks on the Platform for Each Engagement Subtype

Tool	Class 1 OR (LCL-UCL)	P value	Class 2 OR (LCL-UCL)	P value	Class 3 OR (LCL-UCL)	P value	Class 4 OR (LCL-UCL)	P value	Class 5 OR (LCL-UCL)	P value
What's Your Lens? What's Your Thinking Style?	0.97 (0.81-1.16)	.71	0.85 (0.70-1.01)	.07	0.98 (0.85-1.13)	.79	1.28 (1.02-1.60)	.03	1.06 (0.89-1.27)	.51
Mindfulness	1.07 (0.83-1.38)	.62	0.58 (0.43-0.77)	<.001	0.47 (0.38-0.59)	<.001	1.85 (1.39-2.47)	<.001	2.16 (1.72-2.70)	<.001
My TFB Cycles	0.66 (0.56-0.77)[a]	<.001	1.07 (0.92-1.25)	.39	1.31 (1.16-1.47)[a]	<.001	0.84 (0.68-1.02)	.08	0.94 (0.81-1.10)	.47
Activity Scheduling	0.57 (0.48-0.68)[a]	<.001	0.98 (0.84-1.16)	.85	1.32 (1.16-1.50)[a]	<.001	0.95 (0.77-1.18)	.67	1.03 (0.87-1.21)	.74
Challenge Thoughts	0.93 (0.78-1.11)	.43	0.95 (0.80-1.13)	.55	1.20 (1.05-1.38)[a]	.007	0.82 (0.65-1.03)	.09	0.93 (0.78-1.10)	.39
Boosting Behavior	0.67 (0.57-0.79)[a]	<.001	1.01 (0.86-1.18)	.93	1.26 (1.11-1.43)[a]	<.001	0.90 (0.73-1.11)	.33	1.01 (0.86-1.19)	.87
Activities List	0.58 (0.49-0.69)	<.001	1.00 (0.85-1.18)	.99	1.31 (1.16-1.49)[a]	<.001	0.91 (0.74-1.13)	.41	1.03 (0.88-1.21)	.70
Depression: Myths and Facts	0.87 (0.63-1.18)	.37	1.31 (0.99-1.74)	.06	1.08 (0.85-1.36)	.53	0.44 (0.27-0.74)[a]	.002	1.06 (0.80-1.42)	.69
Mood Monitor	1.23 (1.05-1.45)[a]	.011	0.79 (0.67-0.93)	.004	0.51 (0.45-0.59)	<.001	2.81 (2.29-3.45)	<.001	1.04 (0.89-1.22)	.60
Understanding My Situation	1.22 (1.04-1.42)[a]	.01	1.30 (1.12-1.52)	.001	0.94 (0.83-1.06)	.30	0.76 (0.63-0.93)	.008	0.77 (0.66-0.90)	.001
Core Beliefs Quiz	0.94 (0.58-1.53)	.81	0.80 (0.49-1.32)	.39	0.88 (0.61-1.28)	.51	1.93 (1.19-3.16)[a]	.008	0.95 (0.60-1.52)	.84
Hierarchy of Fears	1.09 (0.73-1.62)	.69	0.56 (0.34-0.91)[a]	.02	0.98 (0.71-1.34)	.88	1.33 (0.83-2.15)	.24	1.25 (0.86-1.82)	.25
Anxiety: Myths and Facts	0.84 (0.63-1.14)	.26	1.67 (1.30-2.16)[a]	<.001	1.10 (0.88-1.37)	.41	0.40 (0.24-0.66)[a]	<.001	0.79 (0.59-1.06)	.12
My Worries	1.34 (1.12-1.61)[a]	.002	0.86 (0.71-1.04)	.12	0.96 (0.83-1.11)	.55	1.10 (0.87-1.39)	.44	0.85 (0.71-1.02)	.09
Anxious Thoughts and Worry Quiz	1.45 (1.21-1.74)[a]	<.001	0.90 (0.75-1.08)	.26	0.90 (0.78-1.05)	.18	1.08 (0.85-1.37)	.51	0.82 (0.68-0.99)[a]	.04
Work Stress	0.81 (0.51-1.29)	.37	1.33 (0.89-1.99)	.16	0.93 (0.66-1.31)	.68	0.97 (0.55-1.71)	.92	1.01 (0.66-1.54)	.96
Staying Well	1.08 (0.85-1.36)	.53	0.80 (0.63-1.03)	.08	1.22 (1.02-1.47)[a]	.03	1.21 (0.90-1.62)	.21	0.70 (0.55-0.90)[a]	.005
Activity Goals	1.43 (0.90-2.27)	.13	0.54 (0.29-1.01)	.05	0.39 (0.24-0.64)[a]	<.001	4.51 (2.91-6.98)[a]	<.001	0.83 (0.49-1.39)	.47
Worry Tree	1.26 (1.04-1.52)[a]	.02	0.80 (0.66-0.98)[a]	.03	1.02 (0.88-1.19)	.79	1.11 (0.87-1.42)	.39	0.87 (0.72-1.05)	.15
Graded Exposure Quiz	1.24 (0.86-1.79)	.26	0.59 (0.38-0.93)[a]	.02	0.85 (0.63-1.16)	.31	1.62 (1.06-2.48)[a]	.03	1.13 (0.78-1.62)	.53
Staying in the Present	0.90 (0.77-1.05)	.17	1.17 (1.00-1.36)[a]	.04	1.07 (0.95-1.20)	.30	0.87 (0.71-1.06)	.17	0.91 (0.79-1.06)	.24
Sleeping Tips	0.89 (0.68-1.17)	.40	0.76 (0.58-1.01)	.05	0.91 (0.74-1.11)	.34	1.12 (0.81-1.55)	.49	1.47 (1.16-1.86)[a]	.002
Relaxation	0.50 (0.32-0.78)[a]	.002	0.72 (0.49-1.07)	.10	0.96 (0.73-1.27)	.78	0.88 (0.54-1.42)	.60	2.12 (1.57-2.86)[a]	<.001

Abbreviations: LCL, lower confidence limit; OR, odds ratio; TFB, Thoughts, Feelings, and Behaviours; UCL, upper confidence limit.

[a] Unique significant association.

Table 2. Estimated Engagement Class-Specific Mean Change Over Time

Engagement subtype	Estimated initial score, mean (LCL to UCL) PHQ-9	GAD-7	Estimated change after 14 weeks, mean (LCL to UCL) PHQ-9	GAD-7	Cohen d effect size for change (SE) PHQ-9	PHQ-9
Baseline: class 1	13.35 (13.27-13.44)	12.09 (12.01-12.17)	−5.03 (−4.79 to −5.26)	−4.72 (−4.49 to −4.93)	Baseline	Baseline
Class 2	12.43 (12.21-12.65)	11.53 (11.33-11.73)	−4.41 (−3.91 to −4.90)	−4.18 (−3.71 to −4.63)	−0.46 (0.014)	−0.49 (0.012)
Class 3	12.79 (12.58-13.01)	11.8 (11.61-12.00)	−6.65 (−6.13 to −7.15)	−6.36 (−5.88 to −6.82)	−0.46 (0.014)	−0.48 (0.015)
Class 4	12.96 (12.66-13.27)	11.84 (11.57-12.12)	−5.39 (−4.73 to −5.94)	−4.98 (−4.47 to −5.47)	−0.61 (0.021)	−0.63 (0.022)
Class 5	12.51 (12.26-12.84)	11.61 (11.38-11.84)	−5.88 (−4.84 to −6.38)	−5.56 (−5.08 to −6.02)	−0.73 (0.018)	−0.78 (0.020)

Abbreviations: GAD-7, Generalized Anxiety Disorder-7; LCL, lower confidence limit; PHQ-9, Patient Health Questionnaire-9; UCL, upper confidence limit.

compared with patients in class 5 (estimated PHQ-9 score reduction, 5.88 [0.14]) and class 4 (estimated PHQ-9 score reduction, 5.39 [0.14]). Patients in class 2 had the lowest initial PHQ-9 score and the lowest rate of improvement compared with all other groups, with an estimated mean (SE) improvement of −4.41 (0.13). Class 2 showed the slowest improvement rate, with the lowest mean initial PHQ-9 score (mean PHQ-9, 12.43 [0.07]). Compared with the change in PHQ-9 score in class 1, the Cohen d effect size (SE) was statistically significant at −0.46 (0.014) for class 2, −0.46 (0.014) for class 3, −0.61 (0.021) for class 4, and −0.73 (0.018) for class 5. Similar patterns were found for GAD-7 (Table 2), with class 3 showing the biggest estimated improvement in GAD-7 score after 14 weeks (estimated mean decrease of 6.36 [0.13]), and class 2 showing the lowest improvement (−4.18 [0.52]).

Discussion

In this study, we used probabilistic graphical modeling to identify heterogeneous subtypes of patient engagement with iCBT for symptoms of depression and anxiety. We identified 5 distinct subtypes of users based on program use over 14 weeks. The use patterns of these subtypes suggest that clinical outcomes obtained from interactions with treatment were not always proportional to time spent on the program. Class 4 engaged more in goal-based activities and mood tracking and accessed many core modules, whereas class 5 participants were less likely to access core modules, but used relaxation and mindfulness tools. Patients in class 3 were more likely to complete content belonging to key components of CBT (ie, cognitive restructuring and behavioral activation) within the first 2 weeks on the program. These insights may facilitate tailoring of interventions for specific subtypes of engagement. For example, we may be able to front-load specific recommendations of content associated with improved therapy engagement and clinical outcomes for patients within particular subtypes. Such patterns may elucidate different modalities of engagement that can help us to better triage patients for different therapy modules or activities.

The observed changes in PHQ-9 and GAD-7 scores suggest a dodo bird verdict, where all classes of engagement have won and therefore all show some level of clinical improvement.[29] Thus, even low engagers show a level of engagement that leads to positive outcomes.[14,16] However, reliable clinical improvement for the low engagers was less than noted for those who engaged more with the program and its treatment components. We observed an incremental increase in reliable improvement for patients across the different classes from class 1 (39.5%) to class 5 (66.9%). This increase suggests that offering more personalized interventions may ameliorate these observed differences in clinical improvement rates and increase the possibility of reliable improvement for all.[30,31]

Previous research has focused on the association between simple metrics of mental health intervention use and outcomes. Raw counts of use can serve as a first step toward exploring user behaviors within these programs but they are not able to account for the complexity and diversity in the type of content.[32-34] As a result of the limited perspective of these associations, there has been a call for more sophisticated analytical approaches to better understand how intervention use is associated with clinical outcomes[18,21] and it motivated our present study to assess whether there are groups more likely to respond to the intervention and whether these groups differ in use.[14] A strength of our study is that it included what is, to our knowledge, the largest real-world patient data set of its kind based on over 3 million data points from 54 604 patients from mental health services. Our results suggest that engagement based on iCBT section use in each week gave us a better fit than using engagement as a single binary measure at each week (whether a patient used the program in a given week). Therefore, effective engagement may not be determined merely by absolute engagement with the program, but also by what particular sections or elements a patient engages with. This in turn supports the idea of active treatment ingredients,[34] that is, the components of any treatment that have been empirically supported and may affect therapeutic change. It is precisely these active components that combine to create a coherent treatment.[35]

Future research will need to identify how these machine learning–generated insights about different types of patient engagement and outcome trajectories translate into actionable steps for personalizing the content and delivery of online therapy programs. In this regard, we suggest that the human supporters of internet-delivered interventions, whose role is to provide patients regularly with feedback, continued encouragement, and content recommendations, are well positioned to be the recipients of generated machine learning outputs, providing them with additional information about their patients' engagement and progress with the treatment. As part of their existing practices, these supporters already review patients' activities weekly via platform use metrics (ie, number of logins, page views, and clinical scores), as well as notes that patients enter into their journal or are sent directly to the supporters. Supporters may also benefit from machine learning insights about the behavioral patterns and predicted outcomes for their patients to make more-informed choices in how they personalize treatment recommendations and what they choose to communicate in their feedback. Simultaneously, having a human-in-the-loop and domain expert to evaluate how system-generated insights are to be interpreted foregrounds human agency and leaves the supporters in control and accountable for decisions made on the back of machine learning insights,[36,37] thus presenting a responsible avenue for introducing machine learning into sensitive domains, such as mental health care. Future research at the intersections of machine learning, interaction design, and mental health requires care in identifying how additional subtype information can best fit within existing information review practices for enabling supporters to provide more personalized, effective feedback.

Limitations

This study has limitations. One limitation is that machine learning cannot determine whether there are distinct associations in engagement across sex, different times of year (seasonality effects), or other sociodemographic characteristics as this information is not present in this deidentified data set. Although we are basing our discussion on a robust data set and methods, we have been conservative in our conclusions for clinical implications for iCBT service delivery. This would require further context and replication of the methods on distinct data sets with sociodemographic and clinical covariates to achieve a more complete understanding of the type of profile being considered and what its clinical relevance for iCBT treatment provision might be. These data may help to establish priorities for treatment delivery based on individual needs that promote positive engagement and patient behavior toward realizing improved clinical outcome, and thereby can support any underlying mechanisms that are not ordinarily or immediately observable without machine learning approaches.

Conclusions

In this study, we have defined a path toward triaging patients within an early stage in the intervention to possibly determine whether modifications or additions to the intervention need to take place within or outside the treatment program for maximizing therapeutic benefits. The identification of engagement subtypes may create opportunities for improved intervention strategies, with more tailored, personalized approaches to the delivery of content. By identifying subtypes of patients as opposed to individual level responses, we gain insights regarding typical groups of heterogeneous behavior. Identifying such heterogeneous subtypes may enable us to target distinct groups in a more meaningful way by providing different levels of support or additional treatment content modules.

Author Affiliations: Microsoft Research Cambridge, Cambridge, United Kingdom (Chien, Regan, Carter, Nori, Thieme, Belgrave); Clinical Research & Innovation, SilverCloud Health, Dublin, Ireland (Enrique, Palacios, Keegan, Richards, Doherty); E-Mental Health Group, School of Psychology, Trinity College Dublin, Dublin, Ireland (Enrique, Palacios, Richards); Faculty of Computer Science The University of Vienna, Vienna, Austria (Tschiatschek); School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland (Doherty).

REFERENCES

1. World Health Organization (WHO). *A State of Complete Physical Mental and Social Well-being and Not Merely the Absence of Disease or Infirmity: Constitution of the World Health Organization Basic Documents*. 45th ed. Supplement; 2006.

2. Kessler RC, Angermeyer M, Anthony JC, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*. 2007;6(3): 168-176.

3. Kessler RC, Heeringa S, Lakoma MD, et al. Individual and societal effects of mental disorders on earnings in the United States: results from the national comorbidity survey replication. *Am J Psychiatry*. 2008;165(6):703-711. doi:10.1176/appi.ajp.2008.08010126

4. Alonso J, Angermeyer MC, Bernert S, et al; ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand Suppl*. 2004;109(420):38-46.

5. Substance Abuse and Mental Health Services Administration. *Key Substance Use and Mental Health Indicators in the United States: Results From the 2016 National Survey on Drug Use and Health*. Center for Behavioral Health Statistics and Quality; Substance Abuse and Mental Health Services Administration; 2017. HHS Publication No. SMA 17-5044, NSDUH Series H-52.

6. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013;382(9904):1575-1586. doi:10.1016/S0140-6736(13)61611-6

7. World Health Organization. Fact sheets—depression. Updated January 30, 2020. Accessed June 7, 2020. https://www.who.int/news-room/fact-sheets/detail/depression

8. Richards D, Enrique E, Palacios J. *Internet-Delivered Cognitive Behaviour Therapy—The Handbook of Brief Therapies: A Practical Guide*. Sage; 2019.

9. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. *World Psychiatry*. 2019;18(1):20-28. doi:10.1002/wps.20610

10. Andrews G, Basu A, Cuijpers P, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. *J Anxiety Disord*. 2018;55:70-78. doi:10.1016/j.janxdis.2018.01.001

11. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin Psychol Rev*. 2012;32(4):329-342. doi:10.1016/j.cpr.2012.02.004

12. Wright JH, Owen JJ, Richards D, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. *J Clin Psychiatry*. 2019;80(2):18r12188. doi:10.4088/JCP.18r12188

13. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn Behav Ther*. 2018;47(1):1-18. doi:10.1080/16506073.2017.1401115

14. Donkin L, Hickie IB, Christensen H, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. *J Med internet Res*. 2013;15 (10):e231. doi:10.2196/jmir.2771

15. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. *J Med internet Res*. 2013;15 (8):e172. doi:10.2196/jmir.2258

16. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. *J Med internet Res*. 2014;16(1):e27. doi:10.2196/jmir.2991

17. Fuhr K, Schröder J, Berger T, et al. The association between adherence and outcome in an internet intervention for depression. *J Affect Disord*. 2018;229:443-449. doi:10.1016/j.jad.2017.12.028

18. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. *J Med internet Res*. 2011;13(3):e52. doi:10.2196/jmir.1772

19. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Transl Behav Med*. 2017;7 (2):254-267. doi:10.1007/s13142-016-0453-1

20. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: complex measures of user engagement with an internet-based health promotion intervention. *J Biomed Inform*. 2016;59:299-307. doi:10.1016/j.jbi.2015.12.015

21. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *J Med internet Res*. 2017;19(6):e232. doi:10.2196/jmir.7126

22. Schulam P, Saria S. Integrative analysis using coupled latent variable models for individualizing prognoses. *J Mach Learn Res*. 2016;17(1):8244-8278.

23. Alaa AM, Yoon J, Hu S, van der Schaar M. Personalized risk scoring for critical care prognosis using mixtures of gaussian processes. *IEEE Trans Biomed Eng*. 2018;65(1):207-218. doi:10.1109/TBME.2017.2698602

24. Belgrave DC, Granell R, Simpson A, et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med*. 2014;11(10):e1001748. doi:10.1371/journal.pmed.1001748

25. Richards D, Timulak L, O'Brien E, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. *Behav Res Ther*. 2015;75:20-31. doi:10.1016/j.brat.2015.10.005

26. Richards D, Duffy D, Blackburn B, et al. Digital IAPT: the effectiveness & cost-effectiveness of internet-delivered interventions for depression and anxiety disorders in the Improving Access to Psychological Therapies programme: study protocol for a randomised control trial. *BMC Psychiatry*. 2018;18(1):59. doi:10.1186/s12888-018-1639-5

27. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x

28. Löwe B, Decker O, Müller S, et al. Validation and standardization of the Generalized Anxiety Disorder screener (GAD-7) in the general population. *Med Care*. 2008;46(3):266-274. doi:10.1097/MLR.0b013e318160d093

29. Jacobson NS, Trua P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. In: Kazdin AE, ed. *Methodological Issues & Strategies in Clinical Research*. American Psychological Association; 1992:631-648. doi:10.1037/10109-042

30. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev*. 2006;26(1):17-31. doi:10.1016/j.cpr.2005.07.003

31. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. *J Med internet Res*. 2019;21(8):e12775. doi:10.2196/12775

32. Elliott M, Doane MJ. Stigma management of mental illness: effects of concealment, discrimination, and identification on well-being. *Self Ident*. 2015;14(6):654-674. doi:10.1080/15298868.2015.1053518

33. Chen AT, Wu S, Tomasino KN, Lattie EG, Mohr DC. A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention. *J Biomed Inform*. 2019;94:103187. doi:10.1016/j.jbi.2019.103187

34. Nathan PE, Gorman JM, eds. *A Guide to Treatments That Work*. Oxford University Press; 2015.

35. Duncan BL, Reese RJ. Empirically supported treatments, evidence-based treatments, and evidence-based practice. In: Weiner I, Stricker G, Widiger TA, eds. *Handbook of Psychology*. 2nd ed. Wiley; 2012:489-513.

36. Bellotti V, Edwards K. Intelligibility and accountability: human considerations in context-aware systems. *Hum Comput Interact*. 2001;16(2-4):193-212. doi:10.1207/S15327051HCI16234_05

37. Chikersal P, Belgrave D, Doherty G, et al. Understanding client support strategies to improve clinical outcomes in an online mental health intervention. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Association for Computing Machinery, April 2020. doi:10.1145/3313831.3376341

SUPPLEMENT.
eMethods. Detailed Methods
eTable 1. List of Tools Available on the SilverCloud Health iCBT Platform
eTable 2. Description of Sections Users Engaged With Over a 14 Week Period on the SilverCloud Platform
eTable 3. Estimated Class-Specific Change in PHQ-9 Over Time
eReferences