TBM

Formative, multimethod case studies of learn to quit, an acceptance and commitment therapy smoking cessation app designed for people with serious mental illness

Roger Vilardaga,[1,●] Javier Rizo,[1] Richard K. Ries,[2] Julie A. Kientz,[2] Douglas M. Ziedonis,[3] Kayla Hernandez,[1] Francis J. McClernon[1]

[1]Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA

[2]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

[3]Department of Psychiatry, University of California, San Diego, San Diego, CA, USA

Correspondence to: Roger Vilardaga, roger.vilardaga@duke.edu

Abstract

Despite public health efforts, individuals with serious mental illness (SMI) still have very high rates of tobacco smoking. Innovative approaches to reach this population are needed. These series of case studies aimed to descriptively evaluate the usability, user experience (UX), and user engagement (UE) of Learn to Quit (LTQ), an acceptance and commitment therapy smoking cessation app designed for people with SMI, and to compare it with an app designed for the general population, NCI (National Cancer Institute) QuitGuide (QG). Both apps were combined with nicotine replacement therapy and technical coaching. Inspired by the ORBIT model, we implemented two case studies with crossover AB interventions, two B-phase training designs, and three bi-phasic AB single-case designs with Start-Point and Order randomization (A = QG, B = LTQ). Study outcomes were measured using the System Usability Scale, UX interviews, and background analytics. LTQ's usability levels were above the standard cutoff and on average higher than QG. UX outcomes suggested the relative benefits of LTQ's visual design, gamification and simple design structure. LTQ's overall UE was high; the app was opened for an average of 14 min per day (vs. QG: 7 min). However, users showed low levels of UE with each of the app's tracking feature. Measures of psychiatric functioning suggested the safety of LTQ in people with SMI. LTQ appears to be a usable and engaging smoking cessation app in people with SMI. An optimized version of LTQ should be tested in a Phase II study.

Keywords

Smoking cessation, mHealth, User experience, Case studies, Serious mental illness, Acceptance and commitment therapy

INTRODUCTION

The tobacco smoking rate among adults diagnosed with serious mental illness (SMI), such as schizophrenia spectrum, bipolar, and recurring depressive disorders, is 3 to 4 times the rate of the general population [1, 2]. These high smoking rates have serious health consequences for this population, including a higher incidence of cancer [3] and 25 years of reduced life expectancy [4]. Thus, there is a great need to develop tailored smoking cessation interventions that can be widely disseminated in this population.

Implications

Practice: Learn to Quit may be a viable and safe intervention to increase smoking cessation skills in people with serious mental illness.

Policy: Learn to Quit may be a scalable and cost-effective method of delivery of smoking cessation treatment in people with serious mental illness.

Research: A Phase II study is needed to test the feasibility of this app in a larger sample of this population.

Digital technology may help address the treatment needs of people with SMI. Digital interventions can be standardized to provide evidence-based content, accessed from many locations at any time, tailored to specific groups, and implemented at a lower cost than traditional psychosocial interventions. Furthermore, two recent studies indicate that 72%−81% of individuals with SMI have a mobile device [5, 6], suggesting that wide dissemination of smartphone-based interventions is possible.

Recently, there have been numerous efforts to design digital interventions for people with SMI [7−12]. These include a website to increase psychoeducation [7], two mobile apps developed to improve management of psychotic symptoms [8, 9], and a website for smoking cessation [10, 13]. Despite these efforts, no digital health intervention for smoking cessation has been reported to address smoking cessation in people with SMI using a more ubiquitous and accessible tool such as mobile technology.

Engagement with general smoking cessation apps, however, has been shown to be challenging for people with SMI. This population often experiences deficits in cognitive functioning [14, 15] and theory of mind [16], problems with fine motor skills [17], mental health symptoms [18], and low

educational attainment [19]. In a previous study, we found symptoms of depression and lower education predicted low utilization of SmartQuit, a smoking cessation app developed for the general population [20]. These factors are characteristic of people with SMI, which warranted the need to conduct user-centered design research to identify the design requirements of mobile apps for this population. In a subsequent user-centered design research study [21], we identified a number of critical barriers among people with SMI when using NCI (National Cancer Institute) QuitPal, a smoking cessation app based on U.S. Clinical Practice Guidelines [22]. Two hundred and forty hours of field experience using NCI QuitPal and 10 hr of recorded interviews and task performance revealed (a) considerable guidance needed to complete critical tasks, (b) high task completion latencies ($M = 4.5$ min), and (c) usability levels below recommended standards [21].

Based on this user research, we developed Learn to Quit (LTQ), a smoking cessation app tailored to people with SMI [23]. The app's main active ingredient is acceptance and commitment therapy (ACT) [24] with recommendations from U.S. Clinical Practice Guidelines [22]. LTQ incorporates the following design features to address usability barriers among people with SMI: (a) simple screens, large buttons, and a predictable app structure; (b) gamification of smoking cessation content; (c) use of behavioral principles to enhance retention and comprehension of content; (d) emphasis in visual engagement and storytelling; and (e) access to technical coaching. A detailed report of LTQ's user-centered design research and features can be found elsewhere [23].

The focus of this article is to report the results of a series of formative, multimethod case studies testing the usability, user experience (UX), and user engagement (UE) of LTQ using the ORBIT model for early development of behavioral interventions for chronic conditions as a framework [25]. While our previous research represented Phase Ia of the ORBIT model [20, 21, 23], which aims to define the elements of an intervention, the work reported here is consistent with Phase Ib [25], intended to refine the core elements of the studied intervention in a real-world setting (see Fig. 1).

More specifically, this Phase Ib study allowed us to (a) gain knowledge about the usability of this novel digital intervention in a relatively short period of time without the added costs of recruiting from a larger sample of the population; (b) add confidence in our usability outcomes from a diverse range of measurement approaches; (c) determine whether LTQ's usability and UE were linked to retention and comprehension of the active ingredients of the LTQ intervention—an endpoint that should precede quit rates, our ultimate clinical outcome; and (d) evaluate the acceptability and safety of this type of mHealth intervention in people with SMI.

Case studies are a family of research designs with a wide and flexible range of methodological features [26] including phase comparisons and repeated measures of a relevant outcome [26, 27]. The need for case design methodology has been strongly recommended
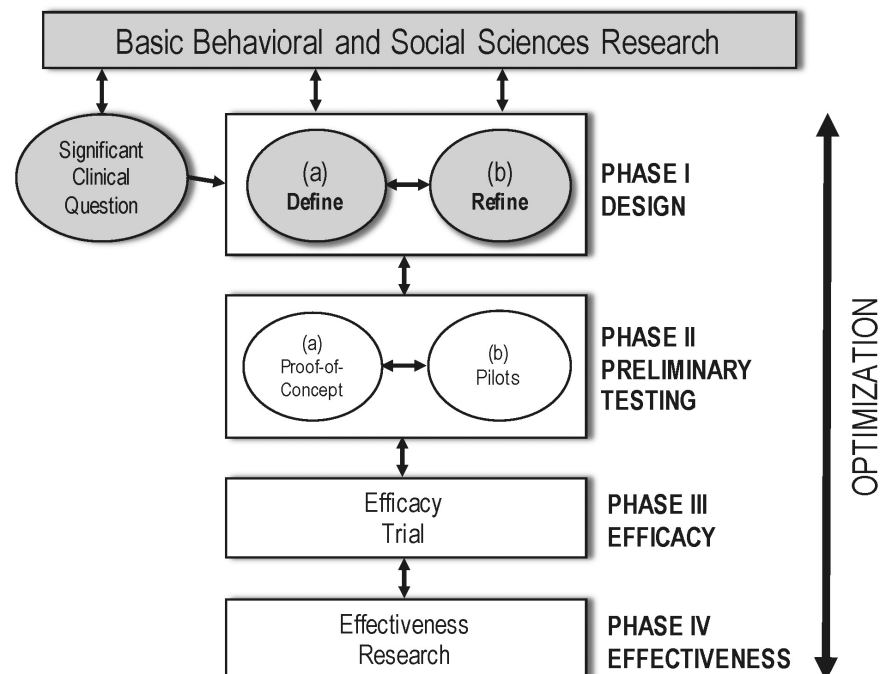

Fig 1 | Learn to Quit's treatment development stage relative to the ORBIT model. Grayed out boxes indicate elements of the ORBIT model addressed by current or past studies.

for early phase treatment development of digital health interventions and as an effective tool to quickly develop treatments in behavioral medicine [25, 26, 28, 29]. Data from these case studies will provide critical support for the need to examine the feasibility and acceptability of this novel intervention in a larger "proof of concept" or pilot Phase II study.

METHODS

Overview of case studies

In a first set of case studies, we evaluated LTQ and NCI QuitGuide (QG) to compare their usability and UX outcomes using a crossover AB design without repeated measures. This design provided an initial evaluation of the usability and UX of each app as measured at the end of each app testing period. In a second set, we examined LTQ's usability, UX, and UE using two B-phase training studies that repeatedly measured UE during an unrestricted 30-day period. This design did not provide a head-to-head comparison of the two apps but offered a more direct test of the natural course of use of a smoking cessation app in a real-world scenario, where the process of quitting often occurs within the first month of setting up a quit date. In a third set of studies, we tested and compared the usability, UX, and UE of both apps using three bi-phasic AB single-case designs. Finally, all case studies examined key individual characteristics that might become barriers to successful use of this novel smoking cessation technology, that is, cognitive performance and psychiatric functioning. Across all case studies, the app intervention was delivered in combination with nicotine replacement therapy (NRT) and technical coaching. The terminology used to describe these case studies is consistent with the Single-Case Reporting Guideline in Behavioral Interventions (i.e., SCRIBE) [30].

One innovation that distinguished our bi-phasic AB single-case designs is that we followed a methodological procedure recently described in the literature [31−33], consisting of the implementation of Order and Start-Point Randomization. Order randomization increases the internal validity of a study by minimizing ordering effects (e.g., LTQ always followed by QG) [32, 33]. Start-Point randomization instead, increases the internal validity of the case study by minimizing length of exposure effects [31, 32]. Start-Point randomization was determined using the Marascuilo-Busk method [34], which we set to a minimum phase length of 7 days per phase using the R Language package SCRT [35]. This randomization scheme produced phase lengths ranging from 7 to 23 days.

Eligibility

We included individuals who (a) were currently receiving treatment at a community mental health clinic; (b) had an International Classification of Diseases, Tenth Revision (ICD-10) diagnosis of schizophrenia, schizoaffective, bipolar, or recurring depressive disorder; (c) self-reported smoking at least five cigarettes per day over the past 30 days, and biochemical verification by expired carbon monoxide test (cutoff: CM > 6 ppm); (d) had a desire to quit smoking within the next 30 days; (e) were 18 years or older; (f) were willing and medically eligible to use NRT; (g) were fluent in spoken and written English; and (h) were taking their psychiatric medications as prescribed by their provider. We excluded individuals who (a) had problematic alcohol or illicit drug use in the last 30 days, (b) had an acute psychotic episode or were unsafe to participate in the study, (c) were pregnant or had the intention to become pregnant in the next 4 months, or (d) were currently receiving any intervention or counseling for smoking cessation.

Procedures

We recruited participants at a local mental health clinic using flyers, provider handouts, and study announcements at the clinic's drop-in center. Diagnostic criteria were verified with the Mini International Neuropsychiatric Interview [36], and problematic alcohol or drug use was screened using the Drug Abuse Screening Test [37] and the Alcohol Use Disorders Identification Test [38]. Smoking status was biochemically verified with exhaled carbon monoxide (CM ≥ 6 ppm) [39]. Adherence to psychiatric medication was self-reported by participant at baseline and follow-up.

Participants who met eligibility criteria took part in a study intake interview at the end of which research staff provided each participant with an android smartphone device (XT1032 Moto G) with access to phone, text, and data. Each device had the randomized app installed according to the Order and Start-Point randomization procedure (Section "Overview of case studies"). Participants were provided with NRT to be used on their quit date. Each week, participants attended a 15−30 min technical coaching session to troubleshoot technical problems and improve their ability to use smartphone technology. For those case studies where there was a phase shift, participants attended an additional session to conduct a semistructured interview to assess their experience with the first app and have the second app installed. At study conclusion, we conducted a second interview, a postassessment of psychiatric functioning, and compensated participants with a $50 gift card.

Interventions

Learn to Quit

LTQ is an Android app developed as a result of formative work that defined key usability barriers and design requirements in smoking cessation apps for people

with SMI [23]. LTQ is based on ACT [24], an intervention that has shown promising results for smoking cessation in a number of clinical trials [40–43], whose active ingredients have shown to predict proximal [44] and distal cessation outcomes [45], and has empirical support as an intervention for SMI [46–48]. LTQ encourages the learning and practice of three processes of change: awareness of urges to smoke, openness to experience urges, and commitment to specific values for quitting. In addition, LTQ adheres to U.S. Clinical Practice Guidelines, including (a) setting up a quit date; (b) preparing for cravings, withdrawal symptoms, slips, and staying smoke free; and (c) use of medications. The app's vision was to design a software infrastructure that would empower users to "learn, practice, and play" skills for quitting, hence its name. LTQ encourages the practice of smoking cessation modules of theory-based content by incentivizing the user with tokens ("Stars"). The app was designed so that it had minimal layers of content presented in successive approximations. Finally, daily ecological momentary assessments (EMAs) were also implemented to track mood, NRT use, cravings to smoke, and number of cigarettes smoked daily. We envisioned this feature as a tool to increase individual's self-awareness of mental health triggers and their relationship with smoking behavior. A more detailed description of the LTQ's app is available elsewhere [23].

NCI QuitGuide
Developed by the NCI, QG is an app based on U.S. Clinical Practice Guidelines for smoking cessation. QG has the following intervention components: (a) psychoeducation about the health consequences of smoking, (b) tracking of smoking habits, and (c) tips for quitting (e.g., getting rid of cigarette ashtrays). More details about the contents and rationale of QG can be found at www.smokefree.gov.

Nicotine replacement therapy
All participants received an 8-week course of NRT (transdermal nicotine patches starting at 21 mg/24 hr) and a 1-week supply of nicotine lozenges (4 mg). This course follows recommendations contained in the U.S. Clinical Practice Guidelines. We instructed participants to use the patches and lozenges on their quit date. A psychiatrist specializing in addictions provided oversight of NRT dispensing and monitoring.

Technical coaching
All participants received smartphone use coaching on a weekly basis by research staff. These coaching sessions lasted between 15 and 30 min. Our coaching intervention had the following components: (a) understanding user familiarity with technology, (b) empathizing with their UX, and (c) taking specific steps toward resolving specific technical issues. We provided technical smartphone consultation over the telephone as needed.

Measures

Baseline characteristics

Positive and Negative Syndrome Scale (PANSS). This 30-item semistructured interview is an extensively used measure of the severity of positive and negative symptoms in psychosis and general psychopathology [49]. Due to the wide range of diagnoses within SMI, we focused on the general psychopathology scale, with scores ranging from 16 to 112 with higher scores indicating more psychopathology, and published norms indicating an average of 44.8 (SD = 9.6) in our target population [50].

Brief Assessment of Cognition. The task is a reliable and valid test of global cognitive functioning that was developed for patients within the SMI spectrum [51]. Following certification by NeuroCogTrials Inc., the task is administered through an iPad app, and examines verbal and working memory, motor speed, verbal fluency, information processing, and executive functioning. A composite score adjusting for age and gender offers a comprehensive view of cognitive functioning, with higher standardized scores indicating better functioning.

Theory of Mind Picture Sequencing Task. [52]. This task assesses individuals' ability to make inferences about others' emotional and mental states, enabling empathic responses and adaptive social communication, a key cognitive factor in SMI. More directly, it assesses our subject's ability to benefit from the LTQ app, which heavily relies on cartoons and storytelling to convey smoking cessation concepts [23]. Higher scores on the False-Belief subtask (Range 0–6) indicate higher theory of mind ability with published norms indicating an average of 3.83 (SD = 1.59) in our target population [53].

Smoking behavior. We collected self-reported years of smoking, smoking levels, and nicotine dependence using the Fagerström Test for Nicotine Dependence [54]. This six-item measure assesses the severity of nicotine dependence. Scores range from 0 to 10, with higher scores reflecting higher dependence. We biochemically verified using Breath-Tests using the piCO+ Smokerlyzer™ with a breath carbon monoxide eligibility cutoff of ≥6 ppm [39].

Usability, user experience, and user engagement metrics

System Usability Scale (SUS). The SUS [55] is a valid and reliable 10-item usability questionnaire widely used by UX researchers. This scale has 10 items with response options on a 5-point Likert scale (1 for "strongly disagree") and a range of total scores ranging from 0 to 100. Scores higher than 68 indicate above standard usability levels [56].

Semi-Structured Interviews (UX). A pool of interview questions was designed to cover the following areas: (a) navigation and design of specific app features, (b) interest in using the app to quit smoking, (c) utility of specific app features, (d) experienced barriers in using the app, (e) situations in which using the app may be pleasant or unpleasant, and (f) overall suggestions for more useful or engaging UX.

Frequency and minutes of app use per day (UE). The "Screen Time" metric in Google Analytics was used to calculate app use frequency and duration for LTQ. This is a conservative metric of app use time because it tracks time of engagement with a specific screen only when activating an "event" (e.g., clicking any element on the screen). Otherwise, time of use of that specific screen is not recorded [57]. QG's frequency and time data was gathered using QualityTime [58], a commercially available app that tracks app's usage. QualityTime was installed in the devices used for the three bi-phasic AB single-case studies. We used app openings at any given time of the day to calculate the percentage of days in which the app was used (e.g., a value of 100% would mean that the app was opened every possible day). Higher percentages indicate higher engagement with the app.

Percent of self-reported EMAs. Google Analytics and QG logs were used to calculate the percentage of use of EMAs features for each app. Higher percentages indicate higher engagement with this tracking feature.

Overall engagement with LTQ's theory-based modules. The number of tokens (i.e., stars) obtained by each subject served as a metric for repeated practice of smoking cessation content. By examining user logs and Google Analytics, we calculated the percent of modules practiced at least three times. For example, a score of 100% would indicate that each of the 28 modules was practiced at least three times, rendering a total of 84 tokens. We examined comprehension and retention of modules' content by analyzing user's responses to built-in LTQ module quizzes (i.e., 42 questions, three per each of the 14 psychoeducational modules). A score of 100% indicates the user responded to all questions correctly.

Data analytic strategy

We used baseline measures (e.g., psychiatric functioning) to contextualize participants' usability metrics in response to known usability barriers in this population. We calculated scores for the SUS, our self-reported measure of usability, at the end of each treatment or phase (A or B) and presented them in each case study figure (Figs. 2–4). These figures reflect the assigned Order (i.e., starting with A or B) and Start-Point randomization.

Upon transcribing the semistructured interviews, two coders (R.V. and K.H.) conducted a formal thematic analysis [59, 60] of these transcripts addressing the following research question: *Do LTQ and QG differ in terms of user experience? In what way are they similar or different?* Our thematic analysis proceeded with the following steps: data familiarization, generation of initial codes, codes collapse, and iterative affinity diagrams [61]. The coders then used affinity diagrams to examine verbal content based on similarity, dependence, and proximity to identify common clusters of content and our final themes. Finally, we extracted UE metrics and LTQ's theory-based
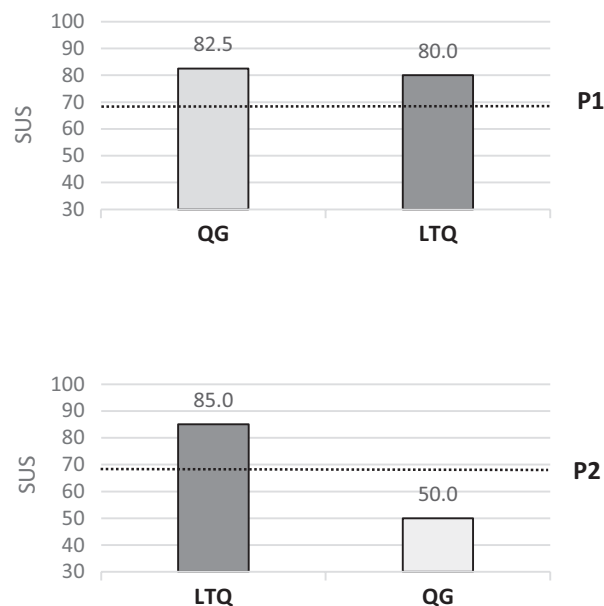

Fig 2 | Case studies with crossover AB interventions. Bars are presented according to cross-over randomization; Dotted lines indicate the 68 cutoff usability standard. *SUS*, System Usability Scale; *P1*, Participant 1; *P2*, Participant 2.
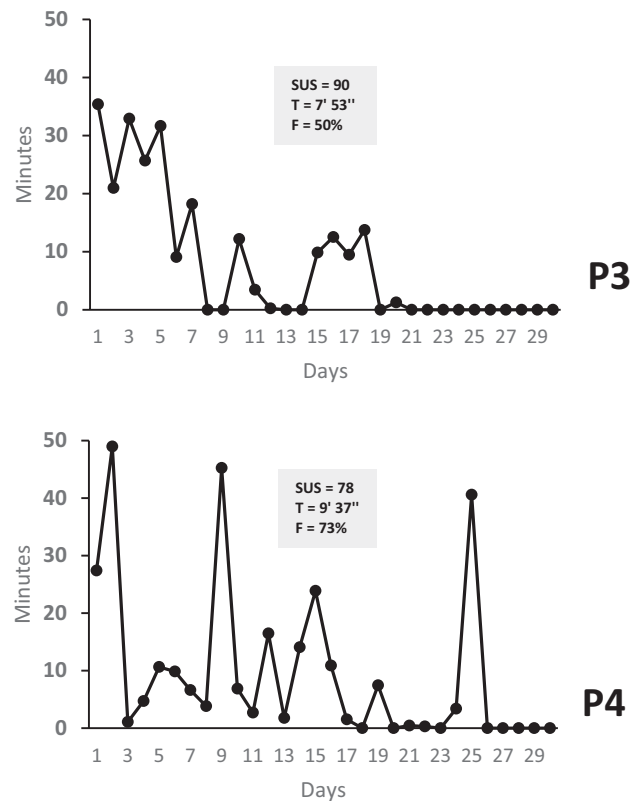
Fig 3 | Results for the B-training studies of LTQ (*n* = 2). *SUS* System Usability Scale; *T* average minutes of use per day; *F* Percent of days in which the app was used relative to the 30-day period; *P3* Participant 3; *P4* Participant 4.

usability metrics from Google Analytics and the QualityTime app.

RESULTS

Participants were recruited from March to May 2016. We screened 38 individuals over the telephone for eligibility in the study. Among those, 14 screened positive for an in-person interview (37%), 9 passed the in-person screen (24%), and 7 completed the study (18%). Dropouts included 1 participant who was hospitalized 3 days after study enrollment due to an unrelated psychiatric event and 1 participant whose study smartphone device was stolen and was unreachable to continue participation.

Average age across participants was 45 (*SD* = 9.5), and four out of seven self-identified as female. Five subjects self-identified as white and two as having more than one race. Four out of seven had high school or less education and an employment disability status.

Case studies with crossover AB interventions

P1 and P2 were two females with psychotic disorders. P1, a patient with a diagnosis of schizophrenia, had deficits in cognitive functioning and theory of mind. Her general psychopathology score was low; however, she presented with significant negative symptoms as reflected by the corresponding scale of the PANSS. P2 was a patient with an unspecified psychotic disorder who had larger deficits in cognitive functioning than P1, as well as lower psychiatric functioning (see Table 1).

LTQ's usability scores were above the standard cutoff (i.e., SUS = 68) in both cases (see Fig. 2), with QG slightly outperforming LTQ in P1, but largely underperforming LTQ in P2. A more detailed thematic analysis of UX interviews (Supplementary Table S1) indicated that P1 had some difficulties reading QG content. She also indicated that she enjoyed the quizzes at the end of LTQ's modules (see [23] for a full description of the app) and found LTQ easy to understand and simple to follow. P2 directly pointed out the ease of use of LTQ compared with QG, stating that her experience using QG was slightly stressful because in her view it felt like completing a chore. However, she commented on new smoking behavior insights gained thanks to QG's tracking features. Both P1 and P2 verbally reframed skills presented by LTQ using their own words, suggesting good retention and comprehension of LTQ content.

UE data indicated that P1 and P2 used LTQ between 10 and 20 min per day, repeatedly practiced LTQ modules, and had high levels of success with LTQ module quizzes. Logs of LTQ and QG in P1 indicated that LTQ's tracking features was
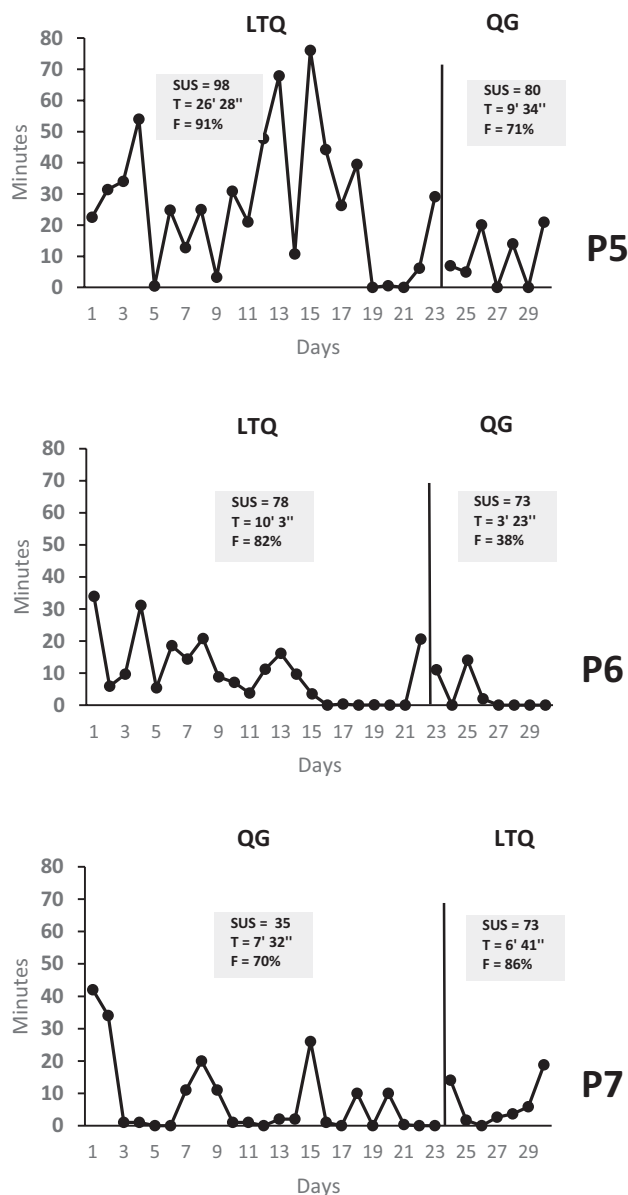
Fig 4 | Results for the bi-phasic AB single-case studies (*n* = 3). *SUS* System Usability Scale; *T* average minutes of use per day; *F* percent of days in which the app was used relative to the 30-day period; *P5* Participant 5; *P6* Participant 6; *P7* Participant 7.

used more frequently than QG tracking features. However, the opposite pattern was observed in P2, which is consistent with her statements during the UX interview.

B-phase training studies

P3 and P4 were a male and a female with psychotic disorders. P3 had a primary mood disorder with psychotic features, and P4 had an unspecified psychotic disorder. P3 presented with very mild cognitive deficits albeit marked deficits in theory of mind. Psychiatrically, he had high levels of general psychopathology. P4 was more psychiatrically stable than P3 but had more deficits in cognition as well as higher nicotine dependence.

These two subjects were only assigned to the LTQ app. Their SUS scores were well above the

usability cutoff, with a very high score in P3. Our interviews indicated very positive UXs in these two subjects (Supplementary Table S1). P3 emphasized the benefits of the interactive features of the app and that he enjoyed its core narrative, which reminded him of a known cartoon character. P4 commented on the emotional impact that LTQ had on her, describing it as a "clean" and "kind" feeling.

UE metrics indicated that averaging across the 30-day period, P3 and P4 used the app at least 50% of all available days with an average that ranged between 8 and 9 min per day. Both subjects practiced LTQ modules several times and responded correctly to all LTQ learning quizzes. Completion of LTQ's tracking feature was low for P3 and significantly higher for P4 (Table 2).

Table 1 | Key baseline characteristics of the sample with relevant pre–post measures indicated

	Primary diagnosis	PANSS[a] Baseline	Post	Years in MH	BAC[b]	ToM-PS[c]	Years smoking	FTND[d]	Carbon monoxide[e] Baseline	Post
P1	Schizophrenia	24	30	25	**-1.38**	**1**	13	4	**14**	**10**
P2	Psychotic disorder	**44**	**41**	26	**-2.16**	5.5	19	4	**9**	5
P3	Mood disorder with PF	**43**	25	35	**-0.01**	5.25	45	3	**19**	27
P4	Psychotic disorder	33	29	28	**-1.09**	6	30	**7**	**19**	4
P5	Mood disorder with PF	**51**	45	10	**-1.78**	2.75	47	5	**13**	10
P6	Mood disorder (Rec)	**44**	26	6	**-0.65**	4.5	16	**6**	**12**	14
P7	Mood disorder with PF	**44**	29	35	0.94	5.5	33	**6**	**9**	2
M	n/a	**40**	32	24	**-.88**	4.35	29	5	**14**	**10**

PF psychotic features; *MH* mental health; *PANSS* Positive and Negative Symptoms Scale–General Psychopathology Scale; *BAC* Brief Assessment of Cognition; *PS-ToM* Picture Sequencing Theory of Mind Task; *FTND* Fagerström Test for Nicotine Dependence.

[a]Bold numbers indicate scores consistent with published norms in this population.

[b]Z scores represent standard deviations from a normative sample matched in age and gender (bold numbers indicate cognitive deficit).

[c]Bold numbers indicate deficits consistent with published norms within this population.

[d]Bold numbers indicate medium or above nicotine dependence.

[e]Pre and 30-days post levels of exhaled carbon monoxide (bold numbers indicate ‹6 ppm cutoff).

Bi-phasic AB single-case studies

P5 and P7 were two individuals with a primary mood disorder with psychotic features and P6 an individual with a recurrent major depressive disorder. All three subjects had levels of psychopathology consistent with normative data in their population. Similarly, P5 and P6 had cognitive and theory of mind deficits consistent with patients in this population. However, P7 performed above average in both the cognitive and the theory of mind tasks. These three subjects were assigned to both the LTQ and QG apps. Across cases, LTQ's usability scores were above the SUS cutoff, and higher for LTQ compared with QG (see Fig. 4). A sharper difference in scores was observed in P7 (38 points). These usability scores were consistent with the results of their UX interviews (Supplementary Table S1). P5 and P6 indicated that LTQ was very easy to use, in P5 despite her report of a diagnosis of dyslexia.

P7 emphasized the benefits of LTQ's interactive and gamification features, and directly commented on the positive effect of LTQ's quizzes on retention of app content. These three subjects also made statements about the challenges they experienced while trying to access specific QG features, including frustration (P5) or simply being surprised during the follow-up interview of some of the key QG features that were available (P6 and P7). P7 enjoyed some of QG's tracking features ("smoke free" tracking button). However, there were inconsistent reports with regard to QG's emotional impact. In one case, QG was criticized for being too "cut and dry" (P7) and

Table 2 | Background analytics for LTQ and QG

	Frequency of use[a] LTQ	QG	Minutes of use per day[b] LTQ	QG	Percent of self-reported EMAs LTQ	QG	Percent modules practiced at least 3 times[c] LTQ	Percent correct quizzes[d] LTQ
P1	91% (10/11)	n/a	9′ 50″	n/a	18% (2/11)	0% (0/19)	74%	96%
P2	100% (11/11)	n/a	20′ 45″	n/a	0% (0/11)	95% (18/19)	90%	100%
P3	50% (15/30)	n/a	7′ 53″	n/a	17% (5/30)	n/a	73%	100%
P4	73% (22/30)	n/a	9′ 37″	n/a	61% (11/18)	n/a	52%	100%
P5	91% (21/23)	71% (5/7)	26′ 28″	9' 34″	26% (6/23)	43% (3/7)	78%	97%
P6	82% (18/22)	38% (3/8)	10′ 3″	3' 23″	23% (5/22)	3/8 (37%)	58%	100%
P7	86% (6/7)	70% (16/23)	6′ 41″	7' 34″	43% (6/7)	52% (12/23)	30%	92%
M	90%	59%[e]	14′ 45″	6'50″[e]	22%	45%[e]	66%	97%

LTQ Learn to Quit; *QG* QuitGuide.

[a]Percentage of assigned days in which the app was opened (raw proportions are indicated in parentheses); [b]Minutes and seconds of app use per day

[c]Percentage of theory-based modules practiced at least 3 times

[d]Percentage of learning module quizzes that were responded correctly.

[e]These averages are based on a smaller number of values; therefore, they need to be interpreted with caution.

in another it was praised for being appropriately "serious" (P6).

UE metrics indicated longer duration of use of LTQ as compared with QG in P5 and P6 (close to a threefold increase), and slightly shorter duration of use of LTQ as compared with QG in P7. Frequency of use was greater for LTQ as compared with QG in all three cases. However, completion of LTQ and QG's tracking features (e.g., cravings, cigarettes) was at or below 52% for both apps. Finally, LTQ-specific metrics indicated that all three subjects practiced LTQ modules repeatedly and responded correctly to almost all LTQ learning module quizzes (see Fig. 4 and Table 2).

Smoking reductions and acceptability and safety of mHealth interventions

Smoking reductions from baseline to study completion are presented in Table 1. Five out of seven participants experienced reductions in smoking, two reported biochemically verified 7-point prevalence abstinence (P4 and P7), and one participant (P2) indicated smoking one to two cigarettes per day, with expired carbon monoxide levels below the cutoff threshold (CM < 6 ppm). Note that we gathered smoking behavior at the end of the 30-day trial period and thus this metric reflects the combined effects of LTQ, QG, and NRT. We evaluated the safety of the mHealth intervention with change scores in the PANSS measure from baseline to the 30-day follow-up. Table 1 shows that psychiatric functioning generally improved or remained stable for all individuals.

DISCUSSION

These seven case studies evaluated the usability, UX, and UE of a smoking cessation app designed for people with SMI. Across case studies, we found that LTQ generally had higher levels of usability, UX, and UE compared with QG, a smoking cessation app developed by the NCI for the general population. UX themes supported LTQ's design, confirmed usability barriers identified in previous research [21], and revealed aspects of QG that were of interest to this population (i.e., tracking and charts). These UX themes also indicated high levels of retention and comprehension of app content, which was consistent with responses to LTQ's quizzes. Our baseline measures confirmed the presence of cognitive deficits in all but one participant, suggesting that the results of these case studies would be generalizable to other patients with SMI and that our efforts to design a smoking cessation app tailored to address these deficits may have been successful. Furthermore, our measures of smoking behavior and psychiatric functioning indicated that mHealth interventions for smoking cessation in this population may be safely delivered without altering patient's psychiatric functioning and lead to positive smoking outcomes.

Objective measures of LTQ's UE indicated that the app was used frequently and for substantial periods of time, and that it led to repeated use of app features linked to processes of change in ACT known to predict cessation outcomes (i.e., awareness of urges, openness to experiencing them, commitment to value-based health actions) [44]. Frequency and app use duration was lower for the B-training studies, which could be interpreted as a result of the fact that LTQ's quit program had a planned duration of 14 days, after which there is no novel content displayed for the user.

Our case studies found some unexpected results. First, in one case (P1), QG's usability was 2.5 points above LTQ's usability, whereas data from the UX interviews indicated that the user found that QG's was difficult to read. We do not have enough information to resolve this discrepancy in metrics; however, given the small difference in usability scores, our interpretation is that on the whole, the subject found both apps equally usable. Second, both apps had low levels of UE with each of the app's tracking features (e.g., cigarette use, mood, craving), with a larger average for the QG app. Although this feature was minimally used in both apps (with the notable exception of P2, who used QG's tracking features 95% of available days), we think these results might indicate that LTQ's EMA design was too dependent upon Android system notifications (i.e., tracking was prompted by the Android system and not linked to a specific app button). In contrast, QG's self-initiated tracking feature (i.e., a button at the center of the app's main screen) was more often used by participants and became one of the themes of our UX interviews. This suggests that when tracking features are incorporated in smoking cessation apps, self-initiated tracking might be a more effective approach to encourage tracking behavior than system-initiated prompts (i.e., notifications).

Based on these findings, we developed an optimized LTQ app that included the following new features: (a) a self-initiated tracking button that the user could access at all times; (b) a wider variety of automated messages in response to self-reported levels of mood or cravings (e.g., "your mood is not too high or too low. Take this chance to practice your skills to quit"); (c) stronger integration of the tracking feature with LTQ modules to increase the personal relevance of the self-tracking feature and increase retention and comprehension of theory-based content. This optimization consisted of adding an automated link at the end of each self-tracking event that would give the user an opportunity to review one of their least practiced modules; (d) an additional set of notifications at the end of the 14-day LTQ journey to encourage long-term review of the least practiced LTQ's smoking cessation modules; and (e) the ability to set a quit date at the end of corresponding quit date module.

Although one participant experienced considerable psychiatric instability during the course of the study, psychiatric functioning generally improved over time for most participants. Our intervention was not designed to reduce psychiatric symptoms. However, given that ACT [24] was originally designed as a mental health intervention, it is possible that some of LTQ's modules might have been useful in addressing their own ongoing mental health symptoms. In fact, four out of seven participants directly stated that LTQ's content was applicable to other areas of their lives. Finally, impaired levels of theory of mind did not seem to have an impact on LTQ's emphasis in storytelling, suggesting that our visual approach to deliver ACT's smoking cessation content was not overly complex for our target population.

The study had several limitations. First, small sample size limits the generalizability of these findings. A larger number of case study replications using the same case method (or a group study) would improve the external validity of these findings. Second, these case studies did not have an experimental control, and the ones that approached it (i.e., bi-phasic AB single-case studies) did not have a sufficient number of replications. While the addition of two randomization procedures (Order and Start-Point randomization) can generally improve the internal validity of a single-case study and bring the design closer to an experimental design, the actual implementation of these procedures led to phases of similar length, not adequately controlling for these biases. To improve their rigor, future studies could use more cases and a new randomization procedure developed by Koehler-Levin [62], which ensures a "staggered" start-point randomization. Despite this, our use of a diverse range of case studies with varying methodological rigor (from case studies without repeated measures to a Bi-phasic AB single-case design) is consistent with the flexibility and focus on agile implementation of this family of research methods [26]. Our methodological approach also used a variety of approaches to measurement of our key target behavior, including time series, single-case designs and qualitative research, all of which are recommended methods to refine interventions in Phase Ib of the ORBIT model that are key to translate novel behavioral technologies into effective health-related treatments [25]. Finally, a larger number of observations in our bi-phasic AB single-case studies would have allowed the use of statistical methods developed to analyze the results of our bi-phasic AB single-case studies [31]. Unfortunately, our 30-day study did not have enough statistical power to conduct such analysis in the same way we have done in previous research [63].

In previous Phase Ia work—as described in the ORBIT model—we identified key elements needed for the design of smoking cessation apps for people with SMI, and the need to rigorously test them [20, 21, 23]. This follow-up Phase Ib study suggests that the resulting theory-based app designed for an SMI population has high levels of usability, UE, and UX. Furthermore, this study indicated appropriate comprehension and retention of LTQ's active ingredients, supporting its promise as a theory-based smoking cessation intervention. Finally, data from this study provided guidance for the development of an optimized LTQ app that could be tested in a Phase II study (e.g., a "proof of concept" study) to evaluate its potential to improve quit rates in our target population.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Translational Behavioral Medicine* online.

Compliance with Ethical Standards

References

1. CDC. Vital signs: current cigarette smoking among adults aged more than 18 years with mental illness—United States, 2009–2011. *MMWR Morbidity and Mortality Weekly Report.* 2013;62.
2. McClave AK, McKnight-Eily LR, Davis SP, Dube SR. Smoking characteristics of adults with selected lifetime mental illnesses: results from the 2007 National Health Interview Survey. *Am J Public Health.* 2010;100(12):2464–2472.
3. McGinty EE, Zhang Y, Guallar E, et al. Cancer incidence in a sample of Maryland residents with serious mental illness. *Psychiatr Serv.* 2012;63(7):714–717.
4. Colton CW, Manderscheid RW. Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. *Prev Chronic Dis.* 2006;3(2):A42.
5. Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: opportunities for future services. *Adm Policy Ment Health.* 2013;40(4):340–343.
6. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of "mHealth" among people with psychosis: a meta-analysis of cross-sectional studies. *Schizophr Bull.* 2016;42(2):448–455.
7. Rotondi AJ, Sinkule J, Haas GL, et al. Designing websites for persons with cognitive deficits: design and usability of a psychoeducational intervention for persons with severe mental illness. *Psychol Serv.* 2007;4(3):202–224.
8. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. *Psychiatr Rehabil J.* 2013;36(4):289–296.
9. Palmier-Claus JE, Rogers A, Ainsworth J, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. *BMC Psychiatry.* 2013;13(1):1–12.

10. Ferron JC, Brunette MF, McHugo GJ, Devitt TS, Martin WM, Drake RE. Developing a quit smoking website that is usable by people with severe mental illnesses. *Psychiatr Rehabil J.* 2011;35(2):111–116.

11. Killikelly C, He Z, Reeder C, Wykes T. Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. *JMIR Mhealth Uhealth.* 2017;5(7):e94.

12. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. *Schizophrenia Bulletin.* 2012;38(3):414–425.

13. Brunette MF, Ferron JC, Robinson D, et al. Brief web-based interventions for young adult smokers with severe mental illnesses: a randomized, controlled pilot study. *Nicotine Tob Res.* 2018;20(10):1206–1214.

14. Green MF, Kern RS, Heaton RK. Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophr Res.* 2004;72(1):41–51.

15. Austin MP, Mitchell P, Goodwin GM. Cognitive deficits in depression: possible implications for functional neuropathology. *Br J Psychiatry.* 2001;178:200–206.

16. Brüne M. "Theory of mind" in schizophrenia: a review of the literature. *Schizophr Bull.* 2005;31(1):21–42.

17. Schwartz BL, Rosse RB, Veazey C, Deutsch SI. Impaired motor skill learning in schizophrenia: implications for corticostriatal dysfunction. *Biol Psychiatry.* 1996;39(4):241–248.

18. Kessler RC, Berglund PA, Bruce ML, et al. The prevalence and correlates of untreated serious mental illness. *Health Serv Res.* 2001;36(6 Pt 1):987–1007.

19. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62(6):617–627.

20. Zeng EY, Vilardaga R, Heffner JL, Mull KE, Bricker JB. Predictors of utilization of a novel smoking cessation smartphone app. *Telemed J E Health.* 2015;21(12):998–1004.

21. Vilardaga R, Rizo J, Kientz JA, McDonell MG, Ries RK, Sobel K. User experience evaluation of a smoking cessation app in people with serious mental illness. *Nicotine Tob Res.* 2016;18(5):1032–1038.

22. 2008 PHS Guideline Update Panel L, Staff. Treating tobacco use and dependence: 2008 update U.S. Public Health Service Clinical Practice Guideline executive summary. *Respiratory Care.* 2008;53(9):1217–1222.

23. Vilardaga R, Rizo J, Zeng E, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. *JMIR Serious Games.* 2018;6(1):e2.

24. Hayes SC, Strosahl KD, Wilson KG. *Acceptance and Commitment Therapy, Second Edition: The Process and Practice of Mindful Change.* New York: Guilford Press; 2011.

25. Czajkowski SM, Powell LH, Adler N, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychol.* 2015;34(10):971–982.

26. Dallery J, Raiff BR. Optimizing behavioral health interventions with single-case designs: from development to dissemination. *Transl Behav Med.* 2014;4(3):290–303.

27. Kratochwill TR, Hitchcock JH, Horner RH, et al. Single-case intervention research design standards. *Remedial and Special Education.* 2013;34(1):26–38.

28. Hekler EB, Klasnja P, Riley WT, et al. Agile science: creating useful products for behavior change in the real world. *Transl Behav Med.* 2016;6(2):317–328.

29. Patrick K, Hekler EB, Estrin D, et al. The pace of technologic change: implications for digital health behavior intervention research. *Am J Prev Med.* 2016;51(5):816–824.

30. Tate RL, Perdices M, Rosenkoetter U, et al. The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement. *Aphasiology.* 2016;30(7):862–876.

31. Heyvaert M, Onghena P. Randomization tests for single-case experiments: state of the art, state of the science, and state of the application. *Journal of Contextual Behavioral Science.* 2014;3(1):51–64.

32. Levin J, Ferron J, Gafurov B. Improved randomization tests for a class of single-case intervention designs. *Journal of Modern Applied Statistical Methods [Internet].* 2014;13(2).

33. Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychol Methods.* 2010;15(2):124–144.

34. Marascuilo LA, Busk PL. Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment.* 1988;10(1):1–28.

35. Bulté I, Onghena P. SCRT: Single-Case Randomization Tests. R package version 1.2.1 [Internet]. 2017.

36. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* 1998;59(Suppl 20):22–33;quiz 34.

37. Skinner HA. The drug abuse screening test. *Addict Behav.* 1982;7(4):363–371.

38. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption–II. *Addiction.* 1993;88(6):791–804.

39. Javors MA, Hatch JP, Lamb RJ. Cut-off levels for breath carbon monoxide as a marker for cigarette smoking. *Addiction.* 2005;100(2):159–167.

40. Bricker JB, Mann SL, Marek PM, Liu J, Peterson AV. Telephone-delivered acceptance and commitment therapy for adult smoking cessation: a feasibility study. *Nicotine Tob Res.* 2010;12(4):454–458.

41. Bricker JB, Mull KE, Kientz JA, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. *Drug Alcohol Depend.* 2014;143:87–94.

42. Gifford EV, Kohlenberg BS, Hayes SC, et al. Does acceptance and relationship focused behavior therapy contribute to bupropion outcomes? A randomized controlled trial of functional analytic psychotherapy and acceptance and commitment therapy for smoking cessation. *Behav Ther.* 2011;42(4):700–715.

43. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. *Nicotine Tob Res.* 2013;15(10):1756–1764.

44. Vilardaga R, Heffner JL, Mercer LD, Bricker JB. Do counselor techniques predict quitting during smoking cessation treatment? A component analysis of telephone-delivered Acceptance and Commitment Therapy. *Behav Res Ther.* 2014;61:89–95.

45. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. *Am J Drug Alcohol Abuse.* 2015;41(1):68–73.

46. Gaudiano BA, Herbert JD. Acute treatment of inpatients with psychotic symptoms using Acceptance and Commitment Therapy: pilot results. *Behav Res Ther.* 2006;44(3):415–437.

47. White R, Gumley A, McTaggart J, et al. A feasibility study of acceptance and commitment therapy for emotional dysfunction following psychosis. *Behav Res Ther.* 2011;49(12):901–907.

48. Bach P, Hayes SC. The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized controlled trial. *J Consult Clin Psychol.* 2002;70(5):1129–1139.

49. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13(2):261–276.

50. Peralta V, Cuesta MJ. Psychometric properties of the positive and negative syndrome scale (PANSS) in schizophrenia. *Psychiatry Res.* 1994;53(1):31–40.

51. Keefe RS, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr Res.* 2004;68(2-3):283–297.

52. Langdon R, Coltheart M. Mentalising, schizotypy, and schizophrenia. *Cognition.* 1999;71(1):43–71.

53. Langdon R, Ward PB, Coltheart M. Reasoning anomalies associated with delusions in schizophrenia. *Schizophr Bull.* 2010;36(2):321–330.

54. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. *Br J Addict.* 1991;86(9):1119–1127.

55. Brooke J. SUS-A quick and dirty usability scale. *Usability evaluation in industry.* 1996;189(194):4–7.

56. Sauro J. A practical guide to the system usability scale: background, benchmarks & best practices. Denver, CO: Measuring Usability LLC. 2011.

57. Google. Session Duration, Avg - Analytics Help [Internet]. Available from: https://support.google.com/analytics/answer/1006253?hl=en. Accessibility verified October 30, 2018.

58. - QualityTime [Internet]. Available from: http://www.qualitytimeapp.com/. Accessibility verified October 30, 2018.

59. Braun V, Clarke V, Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology.* 2006;3(2):77–101.

60. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ.* 2000;320(7227):114–116.

61. IDEO. Human-centered design toolkit: an open-source toolkit to inspire new solutions in the developing world. 2 edition. *S.l.: IDEO.* 2011.

62. Hwang Y, Levin JR, Johnson EW. Pictorial mnemonic-strategy interventions for children with special needs: illustration of a multiply randomized single-case crossover design. *Dev Neurorehabil.* 2018;21(4):223–237.

63. Rosenberg DE, Kadokura E, Morris ME, Renz A, Vilardaga RM. Application of N-of-1 experiments to test the efficacy of inactivity alert features in fitness trackers to increase breaks from sitting in older adults. *Methods Inf Med.* 2017;56(5).

64. Vilardaga R, Rizo J, Zeng EY, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. *JMIR Serious Games.* 2018;6(1):e2. doi:10.2196/games.8881.