Study Protocol and Statistical Analysis Plan (SAP)

DEVELOPMENT OF A COGNITIVE BEHAVIORAL MOBILE APP FOR OBSESSIVE COMPULSIVE DISORDER, AND TESTS OF FEASIBILITY, ACCEPTABILITY, AND EFFICACY

Principal Investigator: **Sabine Wilhelm, PhD**
Chief of Psychology
Mass. General Hospital & Harvard Medical School
Boston, MA, United States of America

Ethical approval: **Mass General Brigham Institutional Review Boards (protocol # 2018P001669, approved on 12/20/2018)**

ClinicalTrials.gov identifier: **NCT04136626**

Funding: **This study was funded by Koa Health. Investigators from MGH developed the Perspectives OCD app in collaboration with technologists and designers from Koa Health via a user-centered, iterative design approach. The investigators from MGH were responsible for the study design and for the execution of the study; authors from Koa Health supported the technical activities related to the development and deployment of the app and provided funding for the project. Koa Health had no role in the recruitment of participants and had no access to data during the trial.**

Document Date: **October 6, 2024 (shortened version)**
Last IRB-approved update: December 6, 2022

Author **Susanne Hoeppner, PhD, MApStat**
Assistant Professor of Psychology
Mass. General Hospital & Harvard Medical School
Boston, MA, United States of America

Version: **Draft 1.0**

Notes: **This version was adapted from a version that included both the development of the Perspectives OCD app (Phase 1) and the randomized controlled trial that tested the app (Phase 2). This version pertains only the protocol sections pertaining to Phase 2.**

DEVELOPMENT OF A COGNITIVE BEHAVIORAL MOBILE APP FOR OBSESSIVE COMPULSIVE DISORDER, AND TESTS OF FEASIBILITY, ACCEPTABILITY, AND EFFICACY

I. BACKGROUND AND SIGNIFICANCE

A. Historical Background

Obsessive compulsive disorder (OCD) is a psychiatric disorder involving unwanted intrusive thoughts coupled with compulsions, or repetitive actions or rituals that are performed to rid oneself of the obsessions (APA 2013). OCD occurs in approximately 2% of the population and was named the 10th leading cause of impairment among all health conditions by the World Health Organization. Unemployment rates in OCD fall between 15-41%, co-occurring depression is experienced by up to 60% of people with OCD, and 11-27% of those with OCD make a suicide attempt. Without treatment, OCD has a chronic and severe course, underscoring the critical importance of access to effective treatment for OCD sufferers.

B. Previous Studies Leading up to the Proposed Research

Cognitive behavioral therapy, and specifically treatment by exposure and response prevention (ERP) therapy, is an effective treatment for obsessive compulsive disorder (OCD) (Abramowitz, Brigidi, & Roche, 2001). Despite the existence of effective treatments, many people do not seek mental health treatment due to perceived stigma about visiting mental health professionals, lack of available mental health providers, and cost of seeking treatment (Mojtabai, 2005; Sareen et al. 2007). Regarding OCD specifically, a majority (57.3%) of individuals with OCD are unable to access treatment, resulting in a major public health concern. Moreover, due to barriers in treatment access, there is an average delay of 10 years between OCD onset and access to care among those who do obtain treatment. Barriers to treatment for OCD include limited availability of professionals who provide this specialized treatment, high costs of in-person treatment, and stigma associated with seeking in-person mental healthcare.

Technology-based interventions such as smartphone and mobile based treatments offer a promising solution to the OCD treatment gap, by providing a low-cost, low-stigma, and widely accessible treatment option. Roughly three-quarters (~77%) of people in the U.S now own a Smartphone (Smith, 2017). Given the ubiquity of Smartphones coupled with the pressing need for highly disseminable mental health treatments, we have developed and are currently pilot-testing a CBT-administered mobile app treatment for body dysmorphic disorder (BDD) (see protocol # 2017P000293). Current preliminary results from our open pilot trial of app-administered CBT for BDD indicate that 90% of patients were treatment responders after completing the 12-week treatment, defined by ≥30% reduction on the BDD-YBOCS. Feedback from our pilot participants indicates that app-based CBT for BDD is feasible and acceptable, and no serious adverse events related to participation have been reported. These preliminary findings suggest that mobile-based CBT treatments may be feasible, safe, and effective. Our CBT for OCD app will be based on very similar cognitive-behavioral principles to those that are used in our CBT for BDD mobile treatment, and our CBT for BDD mobile treatment will serve as the starting point to be adapted for OCD in the present project.

C. Rationale for Proposed Research, and Potential Benefits to Participants and/or Society:

Smartphone-based CBT for OCD can address gaps in access to treatment, offering standardized, high-quality, empirically-based interventions in a low-cost and accessible format. Additionally, given the low rate of doctoral-level clinicians specializing in CBT for OCD treatment, it would be advantageous to establish whether mobile app-based CBT can be administered effectively with assistance from trained bachelors-level coaches who are supervised weekly by licensed, doctoral-level psychologists. This would greatly enhance scalability of the treatment.

II. SPECIFIC AIMS

The overarching purpose of this project is to test the feasibility, acceptability, and efficacy of mobile-app-delivered CBT for OCD. We accomplish this through Specific and Exploratory Aims 1-3.

Specific Aim 1: In a well-powered, randomized-controlled trial, we will test the efficacy of the active treatment arm (Perspectives OCD) compared to a supportive attentional control (HealthWatch) in reducing primary (i.e., OCD symptom severity, measured via Y-BOCS) and secondary outcomes (e.g., quality of life, depression, social and occupational functioning) at week 12.

> Primary Hypothesis: We hypothesize that participants receiving app-based CBT for OCD will have greater improvement in Y-BOCS scores than those in the supportive attentional control condition at treatment endpoint (week 12).
>
> Secondary Hypothesis: We hypothesize that participants receiving app-based CBT for OCD will have greater improvement on secondary clinical outcome measures (i.e., depression (QIDS-SR), functional impairment (WSAS), and quality of life (Q-LES-Q)) than those in the supportive attentional control condition at treatment endpoint (week 12).

Exploratory Aim 2: Feasibility and Acceptability

> Hypothesis: We hypothesize that app-delivered CBT for OCD will be feasible (based on drop-out rates) and acceptable (based on descriptive statistics of patient satisfaction, app use, and app feedback as well as qualitative feedback) to individuals with OCD.

Exploratory Aim 3: Maintenance of gains (Y-BOCS, QIDS-SR, Q-LES-Q-SF, WSAS) at follow-up (week 24).

> Hypothesis: There will be no significant difference between the primary and secondary outcomes measured at treatment endpoint (week 12) and end of follow-up (week 24) in the group that participated in the app-CBT condition.

III. SUBJECT SELECTION

Overview:

Massachusetts General Hospital will be responsible for all participant recruitment and enrollment. We collaborated with Koa Health (formerly Telefónica Alpha, our sponsor and app developers) in the development, iteration, testing, and launch of the app.

In this randomized control trial, our target enrollment is 240 participants. We will aim to randomize 120 eligible individuals with primary OCD. Detailed eligibility criteria follow. Additionally, a detailed plan for safety and risk management is described below.

A. Inclusion/Exclusion Criteria:

1. Inclusion criteria
    a. at least 18 years of age
    b. current diagnosis of primary DSM-5 OCD, based on the MINI
    c. currently living in the United States
2. Exclusion criteria
    a. Psychotropic medication changes within 2 months prior to enrollment
        i. Participants taking psychotropic medication must have been on a stable dose for at least 2 months prior to enrollment and not change medication during study period*
    b. Past participation in ≥4 sessions of CBT for OCD
    c. Current severe substance use disorder
    d. Lifetime bipolar disorder or psychosis
    e. Acute, active suicidal ideation as indicated by clinical judgment and/or a score ≥2 on the suicidal ideation subscale of the C-SSRS.
    f. Current severe comorbid major depression, as indicated by clinical judgment and/or a QIDS-SR total score ≥ 21
    g. Current post-traumatic stress disorder (PTSD)
    h. Concurrent psychological treatment*
    i. Current or past use of an app-based CBT for OCD program (e.g. nOCD, GG-OCD)*
    j. Does not own a supported mobile Smartphone with a data plan
    k. Lack of technology literacy that would interfere with ability to engage with smartphone treatment

*Participants will be asked to refrain from making any medication changes or receiving any concurrent psychological treatment (including other OCD focused app-based programs) during the randomized 12 week treatment period and 3-month follow-up, but will not be asked to refrain from changing their medication or receiving psychological treatment between the 3-month follow-up and 12-month follow-up.

B. Source of Subjects and Recruitment Methods:

Overview:
We never identify potential participants through medical records, and we never contact potential participants without their permission to be contacted. If a medical colleague identifies one of his or her patients to be potentially appropriate for this study, we request that the colleague encourage the patient to contact the PI or the study staff directly. Alternatively, the colleague may

ask the patient to give permission to be contacted over the phone by either the PI or the study staff.

We will not exclude participants based upon gender or minority status. We expect the percentage of minority participants to reflect that of the general population (at least 10-12%), given that OCD presents across ethnicities. We will work to increase enrollment of minority participants by posting advertisements in minority communities (including medical centers), community mental health centers, colleges and universities, and other settings.

We plan to enroll up to 240 participants, in order to meet our target of 120 randomized participants. Potential participants may be informed about the study through MBTA advertisements, radio advertisements, OCD and BDD-focused organizations nationally (e.g., International Obsessive Compulsive Disorder Foundation, Association for Behavioral and Cognitive Therapies), by OCD and BDD clinician and research colleagues nationally, MGH clinician and research colleagues nationally, fliers posted in specialty clinics and hospitals, coffee shops, restaurants, laundromats, barber shops, churches, daycares, libraries, newspapers, universities, other public locations, through our program's website and a recruitment website we will create for this study, Partners Clinical Trials, listservs (e.g., MGH Broadcast), on the Internet, search engine platforms, and in online support groups and social media. When possible, we will turn off the function allowing for comments or interactions on study advertisements. Individuals will also be recruited as part of our clinic's general recruitment protocol #2009P-002227. We will also recruit through the recruitment agency Clinical Connection, who will post our advertisement on their website and share it with their e-list. Interested individuals will be referred to the study RA, who will provide more information about the research study and assess preliminary eligibility over the telephone. We will also use a screening questionnaire on REDCap. The link to access this screener will be embedded within our recruitment website. This will help identify individuals who are likely to be more eligible for the study.

IV. SUBJECT ENROLLMENT

A. Method of enrollment, including procedures for patient registration and/or randomization

Overview

Potential participants will be preliminarily screened by the RA, Dr. Greenberg, or Dr. Weingarden over the phone to establish their likely eligibility. During the phone screen the RA may ask permission to send eligible and interested participants a Partners/MGH Authorization for Release of Protected or Privileged Health Information form, which participants may complete with their psychiatrist's or other healthcare provider's contact information and then send directly to the study RA. If permission is obtained, the study clinician may then contact the participant's healthcare provider to verify issues surrounding potential eligibility (e.g., if there is any question/concern about the anticipated stability of the participant's medication), based on information collected by the RA during the phone screen. All participants will be enrolled by MGH. Potential participants will be given as much time as they need to consider participation, prior to providing informed consent. If suicidality is expressed to an RA during the phone screen, the RA will notify a licensed clinician from the study staff immediately. Per our internal

procedures, RAs will keep the suicidal individual on the phone until a licensed study staff clinician is able to assess the individual.

Interested and eligible participants will be invited to complete the baseline assessment with the IE via any Partners-approved, secure, HIPAA-compliant clinical video platform, such as Enterprise Zoom or Virtual Visit, as recruitment is nation-wide. At that time, patients will be informed about the study's purpose and procedures and advised about alternative treatment options. Before eligibility is assessed, the IE, Dr. Greenberg, Dr. Weingarden, or Dr. Wilhelm will obtain electronic informed consent. The informed consent document will be provided to the potential participant electronically through REDCap and they will be asked to select an "I agree" button and provide their signature in REDCap to indicate their consent (see more REDCap information below: "Privacy and Confidentiality" and "Monitoring and Quality Assurance"). The participant will be informed that they may download and save a copy of the consent form for their records. Self-report measures for the baseline assessment will be completed through a secure REDCap link emailed to participants if participants consent and are deemed eligible.

C. Procedures for obtaining informed consent (including timing of consent process)

Before participating in the screening and baseline assessment, patients will be given information about the study via phone and initial eligibility will be assessed. After this initial phone call, the RA will email interested potential participants a pdf copy of the informed consent document for review, prior to the screening/baseline assessment. The potential participants will be asked to abstain from signing the consent form until study procedures are discussed with the IE, or Drs. Wilhelm, Greenberg, or Weingarden, during the baseline screening video call. In cases where a clinical video platform is used, before the baseline assessment appointments, the RA will send participants an appointment confirmation e-mail with detailed instructions for installing and logging onto any Partners-approved, secure, HIPAA-compliant clinical video platform, such as Enterprise Zoom or Virtual Visit (see index). In advance of the appointments, the RA may conduct a brief test call with the subject to ensure that the subject installed the software (Beiwe for both groups, Perspectives OCD for the CBT group), and can access it from their devices. At this test call, the RA will also collect an emergency contact from the participant, so that the IE will have this information on hand during the baseline assessment. At the start of the screening/baseline assessment, before beginning study procedures, the IE, or Drs. Wilhelm, Greenberg, or Weingarden will inform potential participants about the study's purpose and procedures and advise about alternative treatment options. If the individual wishes to participate in the study, the IE or Drs. Wilhelm, Greenberg, or Weingarden will obtain informed consent electronically by asking participants to click an "I agree" button and electronically sign the electronic informed consent document sent through REDCap, a secure data capture system. The person obtaining informed consent will also verify the participants identity by requesting to view a form of identification (e.g., government-issued driver's license). Participants will have the ability to download and print the electronic informed consent document, or save the pdf copy to their computer for their records. The research assistant will send participants a signed pdf copy of the consent form for their records. The consent process will be documented through checklists in REDCap completed by the IEs and the RA responsible for data entry.

D. Treatment assignment and randomization (if applicable)

Eligible subjects will be randomized to app-based CBT for OCD (Perspectives OCD) or to a supportive attentional control condition (HealthWatch) (50/50 chance). Randomization will be stratified by participants' medication status, in order to evenly distribute any potential medication effects on treatment response across both study arms. Participants in both conditions will be assessed at regular intervals (see Table 2 below). The duration of the supportive attentional control condition will be the same for participants who are treated with CBT. The supportive attentional control group will allow us to determine if active treatment is superior to a psychoeducational website focusing on general health and well-being.

Supportive Attentional Control: To maximize the validity of our study, we will match the duration of the supportive attentional control condition to that of the CBT condition (i.e., approximately 12 weeks). Modules on the supportive attentional control website (HealthWatch) will be made available to participants in parallel to the modules being unlocked on the app in the CBT condition. To protect the safety of participants assigned to the supportive attentional control condition, participants who have active suicidal ideation at the screening assessment (see Inclusion/Exclusion Criteria) will not be eligible to participate. The supportive attentional control website (HealthWatch) provides brief psychoeducational content on OCD and general well-being (e.g., nutrition, heart health), but does not specifically target symptoms of OCD. We will offer referral resources to study candidates who do not wish to be randomly assigned to a possible inactive condition.

STUDY PROCEDURES

Of note, per advisement from Maria Sundquist (Partners IRB) on 2/7/17 (regarding IRB protocol #2017P000293), technologists from Koa Health who will conduct feedback interviews are hired and paid by the sponsor, Koa Health. These interviews are not considered part of the human subjects research and are development work for the app. Therefore, they are not included as study staff in this application, but the informed consent process will fully inform participants about the interactions with technologists and as such, these procedures are described in full in this protocol. MGH study staff will introduce participants to the technology experts via phone call, email, or in person introduction after the participant is fully consented. This will serve to protect participants' identifying information (e.g., we will not share access to participants' email addresses or phone numbers, but rather MGH study staff will initiate conference calls and schedule appointments between participants and technologists).

Email Correspondence
All email communications with participants will be sent in accordance with Partners' Send Secure email encryption policy.

A. Study visits and parameters to be measured

We will conduct a well-powered, randomized trial of the CBT app vs. a supportive attentional control (HealthWatch), to examine its efficacy. We will examine secondary and exploratory outcomes, including treatment feasibility and acceptability (e.g., retention and reasons for treatment refusal and dropout, expectancy, and motivation), and changes in functional

impairment, depression, and quality of life. Therapeutic progress will be broadly assessed with measures of beliefs, behaviors, mood, functioning, and quality of life before, during, and after treatment.

Study Visits and Procedures:

Phone screen: See section IV, "Subject Enrollment," above.

Screening/Assessment Calls and Randomization Procedures: The screening visit will take place over a HIPAA-compliant video conference call (or phone call) and will last approximately 2-3 hours (including an additional hour over the computer for self-report assessments). Subjects will first provide electronic informed consent with the IE, or Drs. Wilhelm, Greenberg, or Weingarden (see Section C, "Procedures for obtaining informed consent" for further details). They will then be assessed by the IE for eligibility, using the below assessment instruments (see Table 2), including a structured diagnostic interview, clinician-rated measures, and self-report measures. Baseline severity and symptom data will also be obtained, using the below assessment instruments. Participants will complete self-report questionnaires via REDCap, online. If an enrolled participant has already completed the same clinician-administered and self-report questionnaires within the past six months as part of a different study within the Center for OCD and Related Disorders, they may be able to consent to give us permission to access data from their previous screening assessment. After the Screening/Baseline visit, participants will be notified by study staff via e-mail (see appendix) regarding their eligibility status, and they will be given the opportunity to discuss this further with our study staff by phone. If eligible, participants will then complete the rest of the self-report tasks and the optional neurocognitive computer tasks (i.e., Stroop Task, Wisconsin Card Sorting Task, below), using the same procedures that were previously approved in Protocol #2017P000293. Ineligible participants will be provided with treatment referrals and resources. When screened patients do not qualify for or choose not to participate in the study, reasons will be documented. Eligible subjects will be randomized through a secure REDCap randomization database in a 1:1 ratio to app-based CBT for OCD or a supportive attentional control condition, stratified by medication status. Those beginning the CBT treatment will be taught to download the app and those beginning the HealthWatch condition will be taught how to access the website. Both groups will receive instruction in downloading the Beiwe app, which was previously approved for use in Protocol # 2019P002041. Please see the schedule of assessments and measures administered at each time point in Table 2. These assessment lengths are similar to those used in other studies in our program and have been well-tolerated by participants. Participants will complete Stroop and Wisconsin Card Sorting neurocognitive computer tasks at home after 4 weeks and at the end of treatment assessment. To avoid dropout, participants will be paid $25 for completing the week 4 neurocognitive tasks, mid-treatment visit, end-treatment visit, and 3-month follow up assessments and $50 for the 12-month follow-up.

Treatment Format: Participants assigned to the CBT condition will complete CBT for OCD through their mobile Perspectives OCD app on their phone over the course of approximately 12 weeks. This app follows the same format as the Perspectives BDD app, which was previously approved for use in Protocol #2017P000293. Treatment length will be 12 weeks. Each treatment component will be presented through modules on the mobile app, and exercises will be logged

and practiced through the app on one's Smartphone. An advantage of app-based CBT for OCD is that participants can self-direct the frequency and duration of modules, and they can re-visit modules as many times as is useful to them. Duration of modules was determined during the development phase (prior study) including considerations based on stakeholder feedback, and will be self-directed by the user based on how long they choose to spend on a given skill.

Upon initiating the app-based CBT treatment, each participant will be assigned a BA-level coach with prior training in CBT and some familiarity with OCD; coaches will be supervised weekly by a licensed psychologist (see "Coach Training and Qualifications," below) and all coach calls will be recorded for fidelity (see "Ongoing Monitoring of Coaching Fidelity" below).The coach will have a brief onboarding phone call to introduce themselves and orient participants to the treatment. Participants can communicate asynchronously with their coach through a secure messaging system incorporated into the app throughout the 12-week treatment. The coach will have a separate administrative interface to the study service which is accessible from inside the Partners firewall and allows the coach to respond to these in-app communications from patients. The coach communication aims to provide support and additional motivational enhancement to patients, and to provide feedback about the skills and homework that the patient is learning through the app. Patients will be notified that the coach will respond to all in-app communications within 36 hours on weekdays. Moreover, brief phone check-ins may be arranged between the participants and coaches on an as-needed basis to supplement the chat system (e.g., to more thoroughly answer a participant question about a skill, help set goals for 2nd half of treatment at treatment mid-point, evaluate and enhance motivation). Participants have access to the Perspectives app and coach support throughout the 12-week treatment. During the initial 3-month follow-up phase of the study, participants will continue to have access to the Perspectives app, but without the coach support or messaging system.

Participants assigned to the control condition will complete the HealthWatch modules through the HealthWatch website over the course of approximately 12 weeks. Upon initiating HealthWatch, each participant will be assigned a BA-level coach with prior training in OCD; coaches will be supervised weekly by a licensed psychologist (see "Coach Training and Qualifications," below) and all coach calls will be recorded for fidelity (see "Ongoing Monitoring of Coaching Fidelity" below). The coach will have a brief onboarding phone call to introduce themselves and orient participants to the intervention. Participants can communicate with their coach via email or telephone throughout the 12-week treatment. The coach communication aims to provide support and additional motivational enhancement to patients. Patients will be notified that the coach will respond to all communications within 36 hours on weekdays. Moreover, brief phone check-ins may be arranged between the participants and coaches on an as-needed basis. Participants have access to the HealthWatch website and coach support throughout the 12-week treatment. During the initial 3-month follow-up phase of the study, participants will continue to have access to the HealthWatch website, but without the coach support.

Coach Training and Qualifications: BA-level coaches will be study staff members with a BA or BS in Psychology or a related field. Based on preliminary findings from our BDD pilot trial (#2017P000293), we anticipate that the coaches' involvement will be light-touch and focus mostly on motivation and problem-solving, as needed, whereas the app/website itself will be the

primary mode of treatment delivery. Specifically, in our pilot trial the coach spent less than 2.5 minutes per patient per week on average responding to chat messages. Research on technology-based treatments likewise suggests that coaches primarily serve to motivate patients and increase accountability and adherence. Coaches will receive training before assisting in treatment as a coach (e.g., completing relevant MGH Psychiatry Academy training course(s)), and will be required to pass (>90% correct) a knowledge test. To ensure ongoing high-quality treatment and to prevent cross-contamination, Dr. Greenberg, Dr. Weingarden, or another study staff member who is a licensed clinical psychologist with expertise in OCD will provide weekly supervision to coaches, with additional supervision as needed if questions arise. Particular care will be taken to prevent cross-contamination (e.g., coaches will be trained and monitored for disallowed strategies, such as discussing specific CBT skills, like exposure, with participants in the HealthWatch condition). All coach calls will be digitally recorded and reviewed at regular intervals by an independent rater for fidelity (see "Ongoing Monitoring of Coaching Fidelity" below).

The Treatment:

The CBT app includes the following components of CBT for OCD: 1) education about a CBT model of OCD; 2) cognitive techniques to identify and challenge distorted thoughts related to one's OCD; 3) exposure to avoided situations and ritual prevention; 4) mindfulness (to help patients to learn to balance distressing emotional states with rational thinking and to control their attentional processes); 5) deeper level (core) beliefs; 6) relapse prevention (to teach patients to expect and react effectively to setbacks that may occur during times of stress). Additionally, passively collected sensor data through the Beiwe app will provide information on changes in participants' location, movement, and phone usage (described below). Our hope is that, in a future stage, the treatment could be adapted to address changes in mobility patterns (e.g., prompting the participant via a message through the app encouraging him or her to engage in the treatment when mobility is notably low).

Assessment and Psychoeducation: The first component of the treatment app will focus on assessing OCD and related symptoms and educating the patient in the CBT model of OCD's development, maintenance, and treatment. Users will set goals for their treatment.

Core Interventions: After assessment and psychoeducation, the app will focus on teaching core cognitive and behavioral methods. For example, cognitive techniques may include skills to help patients identify and evaluate maladaptive OCD-related beliefs. Patients will also learn exposure and response prevention techniques. Patients may also learn mindfulness skills (e.g., observing with a nonjudgmental stance). This approach aims to help patients develop a more accurate view of themselves and the world around them by attending to environmental and social cues other than OCD-relevant ones. Motivational enhancement will be incorporated into the treatment, to help patients increase and maintain motivation for change through treatment. Additionally, patients will learn to identify and evaluate deeper-held (core) beliefs that contribute to their OCD.

Relapse Prevention: The final module will focus on relapse prevention, which aims to help patients maintain their gains after treatment. For example, patients may learn to anticipate

possible symptom recurrence and its relationship to stress, mood, and other factors; differentiate between lapses and relapses; counter negative thoughts about setbacks; and handle lapses and setbacks.

The Supportive Attentional-Control (HealthWatch):
The supportive attentional control is a well-validated control condition, adapted from the website HealthWatch, originally created by Griffiths et al. (2012). The website is comprised of 12 modules related to general health and well-being. Each module contains health information as well as brief questionnaires probing into individual experience with and feelings regarding each topic. The modules provide information on topics such as oral health, cholesterol, and nutrition without alluding directly to anxiety, stress, or depression. For the current study, one of the original modules ("Bacteria & Foodborne Illnesses") was replaced with a brief OCD-specific psychoeducation module. The rationale for replacing this module was two-fold: 1) Contamination is a commonly occurring OCD theme and providing detailed information on bacteria and foodborne illness would likely negatively impact symptoms for at least some participants, and 2) enhancing the control condition with brief OCD-specific psychoeducation and treatment rationale could help boost engagement and credibility. The brief questionnaires are a mix of forced-choice and open-ended questions. The information was sourced from public domain material published by US Government health sources. Each week participants in this condition will be asked to log in to the website, read the information, and fill out the questionnaires assigned for that week. New materials will be presented along approximately the same timeline as in which participants in the CBT arm unlock new modules in the CBT app. As in the CBT-app arm, passively collected sensor data through the Beiwe app will provide information on changes in participants' location, movement, and phone usage (described below).

End-of-Treatment Feedback to Technologists: A subset of participants may meet with technologists from Koa Health (sponsor and Perspectives app developers), either in-person, via audio recorded telephone call, or via any Partners-approved, secure, HIPAA-compliant clinical video platform, such as Enterprise Zoom or Virtual Visit at the end of treatment. The purpose of this assessment is for the technology experts to obtain input and feedback on the usability and feasibility of the CBT app (see "Endpoint technologist interview," attached). This meeting will take up to approximately 1 hour, and participants will be reimbursed $25 for completing this interview.

Measures Descriptions

Diagnostic Measures

Mini International Neuropsychiatric Interview (M.I.N.I.): The M.I.N.I. (Sheehan et al., 2006) is a semi-structured diagnostic assessment of DSM-5 psychiatric illnesses. The M.I.N.I. is efficient, reliable, and well-validated.

Assessment of OCD and Related Symptoms

Yale-Brown Obsessive Compulsive Scale (Y-BOCS): The Yale-Brown Obsessive Compulsive Scale (Y-BOCS); Goodman, Price, Rasmussen, & Mazure, 1989) is a 10-item, clinician-administered measure of OCD symptom severity, which will be our primary outcome

measure. The Y-BOCS assesses individuals' obsessions and compulsions on a scale ranging from 0-4. At the screening visit, clinicians will administer a checklist to preliminary assess which obsessions and compulsions are present.

Obsessive Compulsive Symptoms Rating Scale (OCSRS): The OCSRS (Wilhelm & Steketee, 2006) is a self-report measure of the severity of specific OCD symptom domains.

Clinical Global Impression – Improvement Scale (CGI-I) and severity scale (CGI-S) (Appendix): This rating scale, which ranges from 1 (very much improved) to 7 (very much worse), is commonly used in clinical trials (Guy, 1976). Participants will complete a CGI for OCD symptoms (CGI-P OCD) and the clinician will complete a CGI for OCD symptoms (CGI-OCD) and overall symptoms (CGI-global). The CGI will be a secondary outcome measure and will also be used to determine clinical deterioration of OCD (see Minimizing Risks, below). The CGI also has a severity scale (CGI-S) which is rated by the clinician at baseline. The CGI-S determines the patient's level of severity, in comparison to others the clinician has treated or assessed with the same diagnosis.

Patient Health Questionnaire-2 (PHQ-2) (Appendix): The PHQ-2 (Kroenke & Spitzer, 2002) self-report measure of depression severity includes 2 Likert scale items ranging from 0 (*not at all*) to 3 (*every day*). The 2 items are taken from the longer PHQ-9 measure and are selected because they assess the core diagnostic symptoms of depression. The PHQ-2 will be administered weekly via the app (or REDCap for those in the control condition) to monitor changes in depression symptom severity.3

Quick Inventory of Depressive Symptomatology- Self Report (QIDS-SR) (Appendix): The QIDS-SR (Rush et al., 2003) is a self-report measure of depressive symptoms consisting of 16 scale items with responses ranging from 0 to 3, including one suicide item (item #12). Higher scores correspond with greater depression severity, and the measure is well-validated, sensitive measure of symptom severity in depression. The response choices on the suicide item include 0: "I do not think of suicide or death"; 1: "I feel that life is empty or wonder if it's worth living"; 2: I think of suicide or death several times a week for several minutes"; and 3: "I think of suicide or death several times a day in some detail, or I have made specific plans for suicide or have actually tried to take my life". The QIDS-SR will be used to determine eligibility at the baseline screening visit and to assess depressive symptoms at subsequent assessment visits.

Suicide item (Appendix)*:* A question assessing the presence and nature of suicidal thoughts will be delivered weekly to participants via the app (or REDCap for those in the control condition) to monitor for risk concerns during the trial. Scores >0 will trigger a popup message to the participant about calling 988 or 911/going to the ER and will provide information about contacting a suicide hotline. Scores >1 will also trigger a text message alert to the clinician with the specific item response given by the participant (e.g., 2 vs. 3). The clinician will follow up with the participant within 24 hours by phone to assess for risk, and to refer to a higher level of care if clinically indicated (see Minimizing Risks, below).

Columbia-Suicide Severity Rating Scale (C-SSRS) (Appendix): The C-SSRS (Posner et al., 2010) is a gold-standard, clinician-administered assessment of suicidal ideation and suicide

behaviors. The baseline version, which assesses both lifetime and recent time frames, will be used to establish eligibility at the baseline/screening visit. The "Update" version, which assesses suicide risk since the last assessment, will be used at subsequent study administrations.

Assessment of Functioning and Quality of Life

Work and Social Adjustment Scale (Appendix): The WSAS is a self-report measure of impairment in occupational, social, and family domains (Mundt et al., 2002).

Quality of Life Enjoyment and Satisfaction Questionnaire – Short Form (Q-LES-Q-SF) (Appendix). The Q-LES-Q-SF (Endicott, Nee, Harrison, & Blumenthal, 1993) is a self-report measure of subjective quality of life. Higher scores correspond with greater ratings of quality of life. The Q-LES-Q-SF has strong psychometric properties (Endicott et al., 1993).

Assessment of Expectancy, Motivation and Satisfaction

Expectancy Rating (Appendix): This 4-item self-report questionnaire assesses patients' judgments about the credibility of the treatment rationale, expectancy of change, and treatment acceptability (Borkovec & Nau, 1972). It has good reliability (α=.81-.86), and validity is evident in its ability to differentiate between treatment rationales (Devilly & Borkovec, 2000).

The Client Satisfaction Questionnaire (CSQ) (Appendix) is an 8-item self-report questionnaire which assesses the satisfaction with clinical services received. It has excellent internal consistency and good discriminant validity (Attkisson, 2015).

University of Rhode Island Change Assessment Scale (URICA; Appendix) is a 32-item Likert-type questionnaire that assesses how participants feel about starting and engaging in therapy (e.g., "I am doing something about the problems that have been bothering me.").

Other Assessments

Treatment History Medication, Treatment History Psychosocial, Treatment History Checklist forms (Appendix): These forms will be used to assess any current medications taken as well as any current or lifetime psychosocial treatment.

Lifetime Psychiatric Medication Form (Appendix)*:* This self-report form collects life psychiatric medications taken, and will be administered at the baseline screening visit.

Concomitant medication and therapy form: (Appendix): This log tracks any changes in medication and therapy that the participant has made since the prior assessment.

Demographics Form (Appendix): This self-report form collects basic demographic data and will be administered at baseline.

Treatment Utilization Questionnaire (App/Webstie) (Appendix): This self-report form measures how much time participants are practicing treatment skills both on and off the app or website.

App Feedback Questionnaire (Baseline) (Appendix): This self-report questionnaire asks about participant's expectations of using the app (e.g., "How frequently do you intend to use Perspectives?"

App Feedback Questionnaire (Midpoint) (Appendix): These multiple choice and open-ended questions ask participants to describe their ongoing perceptions of the app at midpoint (e.g., "How easy are the exercises to understand, overall?")

App/Website Feedback Questionnaire (Endpoint) (Appendix): This self-report form collects participant feedback pertaining to the content and aesthetics of the app or website.

Endpoint Technologist Interview Notes (Appendix): This interview guide asks about credibility/expectancy, client satisfaction, and improvements at Week 12 Endpoint

Treatment Condition Questionnaire (Appendix): This brief measure will be used to assess whether the independent evaluator believes each subject was assigned to either CBT or HealthWatch as well as their confidence in this belief. This questionnaire will be completed at the end-of-treatment time point. This measure is based on Bang's 2x5 Blinding Index (Bang et al., 2004).

Life Events Questionnaire (Appendix): This question will be used to assess whether any major life events occurred that might have had a psychological impact on the subject.

COVID-19 Impact Questionnaire (Appendix): This 18 item self-report questionnaire asks participants about the impact of the coronavirus pandemic on their lives (e.g., "The COVID-19 pandemic and its impact have made my symptoms worse."). This measure will be administered to participants for all assessment visits.

Generalized Anxiety Disorder Screener (GAD-7, with COIVD item) (Appendix): The original 7-item self-report questionnaire asks participants about anxiety symptoms experienced in the past two weeks. It has good to excellent internal consistency (Spitzer et al 2006; Loewe et al., 2008) and good convergent validity with other anxiety scales (Spitzer et al., 2006). We added one additional item to ask participants to self-rate how much their anxiety changed relative to before the COVID-19 pandemic. This measure will be administered to participants for all assessment visits.

Substance Use Questionnaire (with COVID item) (Appendix): This 2-item self-report questionnaire with 4 logic-branched follow-up questions asks participants about past month alcohol and other substance use. This measure will be administered to participants for all assessment visits.

Mobile App Rating Scale: User Version (uMARS) (Appendix): This 23-item user-rated scale contains 4 objective quality subscales—engagement, functionality, aesthetics, and information quality—as well as a subjective quality rating. The MARS has shown high levels of interrater reliability for evaluating the quality of mHealth apps (Stoyanov et al, 2015). This measure will be completed at the end-of-treatment time point.

Wi-Fi Questionnaire: This 1-item measure sent weekly via Beiwe assesses whether the participants has connected to Wi-Fi so that study data can be downloaded. This feature is a tech requirement of the Android and iOS platforms.

Self-Help Questionnaire: This question will be used to assess what types of self-help programs an individual has engaged in (e.g. exercise, self-help books or non-OCD apps, peer-support groups)

IE Training and Qualifications, and Procedures to Ensure Assessment Integrity and Interrater Reliability: Assessments will be conducted by an independent evaluator (IE) who has a Masters or Doctoral degree in clinical psychology or related mental health field and will be employed at MGH. The IE will be otherwise uninvolved in study procedures. Training and reliability checks will be done to ensure that IEs conduct ratings in a uniform way. IEs will first receive instruction in the MINI, and Y-BOCS, from Dr. Wilhelm or another gold-standard expert rater, prior to beginning as an IE. The IE will be supervised twice monthly by the PI. The IE will be required to demonstrate reliability on the Y-BOCS at a criterion of .80 ICC and 100% agreement on OCD diagnosis on the MINI, compared to measures rated by Dr. Wilhelm or another gold-standard rater. All assessments will be audiotaped for reliability ratings. Names will not be included on digital recordings. To reduce rater drift, a trained reliability rater meeting the same qualifications as the IE will review 15% of randomly selected audiotaped interviews. If reliability falls below .75, we will institute retraining procedures.

Neurocognitive Computer Tasks

Participants will have the option to complete a set of computer tasks (the Stroop and Wisconsin Card Sorting Tasks), which will be administered via *Inquisit* by Millisecond, an online platform for precision psychological testing. To access the tasks, participants will be provided with a link and unique participant ID to log in. The tasks take about 15 minutes total and will be completed on a computer at home.

Inquisit's system is highly secure and customizable. All files and data are encrypted in transit and at rest, and IP addresses are not collected. No PHI will be collected from participants. The task link is not accessible to the public and is customized for this particular study. Participants will log in through a standard Secure Socket Layer (SSL) technology, which ensures protection of privacy and secure data, via data encryption and server authentication. Data will be transferred and deleted from Millisecond's web-based server after data collection. The database is password-protected and will only be accessible to staff directly involved in the research. Staff members will have unique passwords that comply with PHS policy (minimum 8 characters, alphanumeric, change every 90 days), and unique data access rights. Staff members with rights can delete data from the server at any time, making them inaccessible to Millisecond.

The Stroop Task: This neurocognitive task employs both controlled (color naming) and automatic (reading) processes. Participants are asked to complete the approximately five-minute task by naming the color ink of a word while ignoring the meaning of the word. In the classic Stroop Task, all words are the names of colors. The difference in time between naming the ink of a different color-word (e.g. the word 'red' is written in blue ink) versus naming the ink of a same color-word (e.g. the word 'red' is written in red ink) demonstrates the interference of automatic reading processes on controlled color naming processes (Stroop, 1935). In this computerized version of the task, participants will "name" colors by pressing specific keyboard keys. A larger Stroop interference represents difficulty with selective attention processes, processing speed, and response inhibition (e.g. Ben-David, Nguyen, van Lieshout, 2011).

The Wisconsin Card Sorting Test (WCST): This neurocognitive task measures executive function and strategic planning, particularly with the ability of an individual to use feedback from the environment to shift individual cognitive sets and experience goal-directed behavior. During this 10-15 minutes test, four stimulus cards incorporating color, form, and number are presented to a subject. The subject must sort the cards (i.e. press certain keyboard keys) according to different rules or principles and be able to change their approach throughout the test administration (Grant & Berg, 1981).

Measures Chart OCD

Form Type	**Data Collection Method**	**Measure**	Baseline	Baseline Re-assess	Weekly	Week 4	Mid-Treatment Assess (wk 6)	Post-Treatment Assess (wk 12)	3-month Follow-up	12-month Follow-up
Diagnosis & Screening	Paper/PDF/REDCap	MINI	x							
	REDCap	Demographics	x							
	REDCap	Treatment History	x							
	REDCap	Treatment History	x							
	REDCap	Treatment History	x							
	REDCap	Lifetime Meds	x							
OCD & Related Symptoms	Paper/PDF/REDCap	Y-BOCS	x	x			x	x	x	x
	REDCap	OCSRS	x				x	x	x	x
	REDCap	QIDS-SR	x	x			x	x	x	x
	App/ REDCap	PHQ-2			x					
	App/ REDCap	Suicide item			x					
	REDCap	C-SSRS: Lifetime	x							
	REDCap	C-SSRS: Update		x						
	REDCap	CGI-S: Clinician	x	x						
	REDCap	CGI-I: Clinician		x			x	x	x	x
	App/ REDCap	CGI-I: Patient (OCD)			x			x	x	x
Functioning	REDCap	Q-LES-Q	x				x	x	x	x
	REDCap	WSAS	x				x	x	x	x
Safety	REDCap	CONCOM		x			x	x	x	x
	REDCap	Adverse Events Form		x			x	x	x	x
	REDCap	Life Events		x			x	x	x	x
Tx expectancy, motivation,	REDCap	Cred/Expectancy	x				x			
	REDCap	CSQ-8					x	x		
	REDCap	URICA	x							

Form Type	**Data Collection Method**	**Measure**	Baseline	Baseline Re-assess	Weekly	Week 4	Mid-Treatment Assess (wk 6)	Post-Treatment Assess (wk 12)	3-month Follow-up	12-month Follow-up
satisfaction, feedback	REDCap	TUQ - App†	x				x	x	x	
	REDCap	TUQ – Website*	x				x	x	x	
	REDCap	App Feedback†	x				x	x		
	REDCap	Website Feedback*						x		
	REDCap	Technologist							x	
	REDCap	uMARS†						x		
Neurocognitive	REDCap/Inquist	Stroop	x			x		x		
	REDCap/Inquist	WCST	x			x		x		
COVID	REDCap	COVID Impact	x				x	x	x	x
	REDCap	GAD-7 with COVID	x				x	x	x	x
	REDCap	Substance Use	x				x	x	x	x
Other	Beiwe	Wifi check-in question			x					
	REDCap	Self-Help Questionnaire	x				x	x	x	x

† Only included for participants in the CBT app condition
*** Only included for participants in the HealthWatch condition**

If more than ten days elapse between a patient's initial baseline screening assessment and the start of treatment, the study IE would re-administer select forms, including an additional Y-BOCS and QIDS-SR form.

Despite our best efforts to retain participants, we do expect dropouts. Except for subjects who withdraw consent to participate, all who are withdrawn or drop out of the study will be asked to complete all scheduled assessments and we will provide renumeration for participation in assessments

In addition to the assessment measures in Table 2, de-identified data from sensors in participants' mobile phones will also be collected via the Beiwe app, to inform design and improvement of the app.

Passively Collected Data to be Measured: We will collect de-identified data from sensors in participants' smartphones during the study, using the study Beiwe app. This app was previously approved for use in one of our program research studies; Protocol # 2019P002041. The Beiwe app will collect information about location (using phone GPS), movement (e.g., using phone accelerometer, gyroscope, magnetometer, proximity), and phone usage (power state, phone/screen usage). It may also monitor how much the participant uses their phone for calling and texting and keep track of the people they communicate with (in an anonymized way; please see "Privacy and Confidentiality" below. Note that some of these data streams are available only from iOS or Android, respectively. In order for Beiwe to continue to run in the background of participants' phones across the study period, participants must occasionally interact with the Beiwe app. Data from sensors in participants' mobile phones may also be collected, to optimize the program through personalization and improvement of the app. This may include the following. (For a detailed description of storage and protection of de-identified mobile data, see

Monitoring and Quality Assurance, below.) We will use sensors in the phone to collect data on participant's mobility and app usage.

Quantifying mobility: Obtaining data on participants' mobility offers a useful way to detect clinically relevant changes to a person's mental health status. Therefore, mobility patterns (e.g. the time spent at home) provide highly relevant information to monitor in our app-based treatment. Beiwe will collect location and movement data using the phone GPS, accelerometer, gyroscope, magnetometer, and proximity. For details on how this information will be stored and anonymized please see "Privacy and Confidentiality" below.

Phone Usage Patterns: Dynamics of the phone usage provides an additional insight into clinically relevant changes to a person's mental health status. For that reason, Beiwe will collect phone-unlock events and data traffic to characterize patterns and quantify daily phone and internet usage. Note that the data traffic refers only to the periodically sampled number of bytes sent and received, and it does not include any sensitive information such as internet history or the names of the installed applications. Beiwe may also monitor how much the participant uses their phone for calling and texting and keep track of the people they communicate with. To encourage occasional interaction with the Beiwe app, we will administer a weekly survey question to participants via Beiwe (Wifi check-in question).For details on how this information will be stored and anonymized please see "Privacy and Confidentiality" below.

However, sampling raw GPS location coordinates may intrude on one's privacy. To protect privacy of the study participants, we will apply the following procedure. The location coordinates will be first collected at the secured server located at the hospital or locally on the phone. We will subsequently remove the raw location coordinates by replacing them with randomly generated strings as location labels (such as "ghhu45", "235oh4", "8n8hj3", …), where each unique location in the dataset will correspond to a unique de-identified label. This will allow us to quantify mobility of the participants without requiring us to store actual geographical locations in the long-term. In addition to location, other sensor data collected by the participant's Smartphone may be used to quantify mobility, including accelerometer, steps, calories burnt, and sedentary time.

Application Usage Metrics: To improve the usability of the application, usage metrics may be collected (for example operating system version and device model, time and date when the application is opened or closed, time spent on each page visited, notification timing, etc). Additionally, we may collect user's battery status and charging patterns, and network traffic (e.g., number of bytes sent/received, and the hashed Wi-Fi-antenna indicator). These metrics will be stored in the same fashion as passive sensor data and user survey responses collected via the app.

B. Drugs to be used

Not applicable

C. Devices to be used

Participants assigned to the CBT condition will download the CBT for OCD app onto their personal smartphone devices. The CBT for OCD app is an investigational device in the United States.

D. Procedures, surgical interventions, etc.
Not applicable

E. Data to be collected and when the data will be collected
See above for the assessment schedules and assessment batteries.

VI. BIOSTATISTICAL ANALYSES

The IE and the RAs will play the primary role in data entry. Data will be entered in REDCap, a HIPAA-compliant, Partners-approved platform for electronic data capture that streamlines data collection and management, and ensures data integrity, resulting in improved data quality and reduced costs. For more information on REDCap, please see section IX, "Monitoring and Quality Assurance."

Dr. Susanne Hoeppner will be responsible for data management and analysis. All project staff will receive training in data management and data confidentiality procedures. Data checks will be done regularly to assure that all forms are entered and available for analysis. Questions or problems will be resolved promptly by communication between study staff. Data and analysis files will be backed up on the lab server and may also be stored in separate locked cabinets.

To characterize our sample, data will be displayed graphically, and summary statistics (e.g., means and frequencies) will be calculated for all variables, including demographic and clinical descriptors (e.g., from the MINI).

Baseline differences between the treatment arms in demographic and other potential prognostic variables will be examined using chi-square analyses for discrete variables (if prevalence >10 in the treatment arm with the lower prevalence, or Fisher's exact test if not) and t-tests for continuous variables. Prior to data analysis, all major variables will be screened for inconsistent or abnormal values. Continuous measures will be assessed for skewness and outliers (based on model residuals), and, if needed, will be transformed to better meet modeling assumptions of normality and homogeneity of variance. A two-tailed p-value <.05 will be considered evidence of statistical significance for the primary and secondary outcomes. Our primary and secondary analyses will use generalized linear mixed-effects models (GLMMs). GLMMs can easily accommodate nesting of repeated observations within subjects, include all who complete at least one assessment (including the baseline assessment), and are the preferred method to analyze longitudinal data (Hamer and Simpson, 2009). Despite our best efforts to retain participants, we do expect dropouts (subjects who are lost to follow-up or withdraw early from the study). We will attempt to perform all scheduled assessments for subjects who are withdrawn from the protocol and will provide financial incentives for participation in assessments. Our intent-to-treat sample will include all randomized patients who complete at least one assessment, making our

primary and secondary analyses intent-to-treat analyses. We will also perform "per protocol" analyses. Our per protocol population will include all randomized patients who complete the baseline and post-treatment assessments, and who did not initiate prohibited treatment during the study. We will repeat our GLMM analyses using the per protocol sample. Dropout, study withdrawal, and loss to follow-up will be tabulated by reason and treatment arm. Analyses will be conducted to ascertain to what extent dropouts are nonrandom, and, if so, what factors are associated with dropout. To evaluate potential attrition bias, study dropouts will be compared to study completers.

A. Specific data variables being collected for the study (e.g., data collection sheets)
Specific variables being collected and their timeline are presented above, in Tables 1 and 2, and described in the Study Procedures section.

B. Study endpoints

Primary endpoint: Difference in OCD severity (Y-BOCS) at the end of treatment (i.e., week 12 assessment).

Hypothesis 1: Participants receiving app-CBT will have lower Y-BOCS scores than those in the supportive attentional control condition at treatment endpoint (week 12).

Secondary endpoints: Difference in secondary clinical outcomes at the end of treatment (i.e., week 12 assessment)

Hypothesis 2: Participants who receive the app-CBT will have better scores on secondary clinical outcome measures than those in the supportive attentional control condition at treatment endpoint (week 12) with respect to:

H2.1: depression severity, as measured by the QIDS-SR, where lower scores are better.

H2.2: functional impairment, as measure by the WSAS, where lower scores are better.

H2.3: quality of life, as measured by the Q-LES-Q-SF, where higher scores are better.

Exploratory endpoint 1: Feasibility and Acceptability

Hypothesis 3.1: Drop-out from app-delivered CBT for OCD will not be higher than drop-out from the supportive attentional control condition, which will indicate feasibility.

Hypothesis 3.2: Patient satisfaction ratings (i.e., CSQ total scores) in the group receiving app-delivered CBT for OCD will not be lower than those in the supportive attentional control condition, which will indicated acceptability.

Hypothesis 3.3: App-delivered CBT for OCD will be acceptable to individuals with OCD based on descriptive statistics of app use, app feedback, and qualitative feedback comments.

Exploratory endpoint 2: Maintenance of gains (Y-BOCS, QIDS-SR, WSAS, Q-LES-Q-SF) during follow-up (at 3-month and 12-month follow-up assessments).

Hypothesis 4: At the 12-month follow-up assessment, participants who received app-CBT will have lower Y-BOCS scores than those in the supportive attentional control condition.

Exploratory endpoint 3: Treatment utilization for OCD concerns following the 3-month follow-up.

Hypothesis 5: Participants assigned to the app-CBT treatment condition will be less likely to seek additional treatment for OCD than participants assigned to the supportive attentional control condition prior to the 12-month follow-up assessment.

C. Statistical methods

For all statistical analyses, we will use a Type I error probability of 0.05 to determine significance unless otherwise noted.

Primary endpoint: Difference in OCD severity (Y-BOCS) at the end of treatment (i.e., preliminary efficacy for primary OCD symptoms)

Analysis:

The primary outcome model will be a hierarchical mixed model (i.e., GLMM) that will include time (categorical; baseline, mid-point, and end-of-treatment), treatment (app-CBT vs. supportive attentional control), and their interaction as fixed effects, and will model time as a repeated measure using either an autoregressive (AR1), Toeplitz, compound symmetry, or unstructured covariance matrix, based on best fit determined by AIC and BIC. The unstructured covariance matrix model will only be used if this model fits significantly better than alternative simpler models, based on -2 log likelihoods. The main hypothesis test will be based on a specific contrast of treatment difference at week 12. Because these analyses will include all subjects who are randomized and complete at least one assessment, it is an intent-to-treat analysis. These analyses will include only the 3 assessments from baseline to post-treatment and will not include the 3-month or 12-month follow-up (FU) assessments, because treatment response could be affected by subjects seeking other treatments during the FU period. Between-group effect sizes will be calculated using Cohen's *d*. This analysis will then be repeated using the per-protocol sample.

In addition to the analyses specified above, we will describe the proportion of subjects in each treatment arm who achieve response, where response is defined as a 25% or greater reduction in Y-BOCS score from baseline to end of treatment (week 12) (Storch et al., 2006). While no formal comparisons will be made between treatments, the reporting of response rates is clinically useful and will facilitate comparison to other studies in the field. We will report response as observed (excluding subjects with missing data at week 12) and will additionally examine how rates may change when a) applying a last observation carried forward approach, and b) conservatively considering subjects with missing week 12 scores as non-responders.

Secondary endpoint: Difference in secondary clinical outcomes (i.e., QIDS-SR, WSAS, Q-LES-Q-SF) at the end of treatment (i.e., preliminary efficacy for secondary symptoms)

Analysis:
We will use the same hierarchical mixed modeling approach as described for the primary endpoint to examine significant differences in secondary outcome measures (i.e., depression severity (H2.1), functional impairment (H2.2), and quality of life (H2.3)) at post-treatment between participant in the app-CBT vs. supportive attentional control conditions. We will not adjust for multiple testing among secondary outcomes based on the recommendation of Cook and Farewell (1996), who argue that multiplicity adjustments are not necessary if separate test results are interpreted marginally and address different aspects of the patient experience and decision-making process rather than alternative assessments of efficacy.

Exploratory endpoint 1: Assess feasibility and acceptability

Analysis:

We will test feasibility and acceptability by examining:

H3.1: Refusal and dropout rates and reasons, by group: Drop-out rates by treatment (app-CBT vs. supportive attentional control) will be compared using logistic regression. Drop-outs will be defined operationally as participants who do not complete either an endpoint or follow-up assessment of the primary outcome measures (Y-BOCS), as well as participants in the CBT arm who do not use the app or participants in the HealthWatch arm who do not use the website at least once between assessments (i.e., at least once between baseline and week 6 and at least once between week 6 and week 12). We will use one-tailed tests to examine if the odds of drop-out in the app-CBT group are higher; if non-significant, we will interpret the results as an indication that the app-CBT treatment is feasible in comparison to the supportive attentional control condition; conversely, significantly higher odds of drop-out in the app-CBT condition would indicate a lack of feasibility.

H3.2: We will analyze patient satisfaction, as measured by on Client Satisfaction Questionnaire (CSQ) in a hierarchical mixed model (i.e., GLMM) that will include time (categorical; week 6 and week 12), treatment (app-CBT vs. supportive attentional control), and their interaction as fixed effects, and will model time as a repeated measure using a compound symmetry covariance structure. We will test group differences with one-tailed tests in patient satisfaction (testing for lower satisfaction) at both week 6 and week 12 using specific contrasts with α=.025 to adjust for multiple testing. Non-significant results of these tests will be interpreted as indications of treatment acceptability, while lower satisfaction will be interpreted as indicating a lack of acceptability.

H3.3: Quantitative (i.e., app feedback questionnaires, uMARS) and qualitative participant feedback about app features will be summarized for the app-CBT group using descriptive statistics.

Exploratory endpoint 2: Maintenance of gains (Y-BOCS, QIDS-SR, WSAS, Q-LES-Q-SF) at follow-up (week 24).

Analysis:

We will use hierarchical mixed models similar to the ones used for the primary and secondary outcome analyses, but will add follow-up assessments (i.e., 3-month and 12-month) to the data as additional time-points. We will then use specific contrasts (for each outcome model) to test for significant treatment group differences at the 12-month follow-up time-point. In the case that at least 10% of the participants who completed the end-of-treatment assessment initiated additional or subsequent evidence-based treatment for OCD at clinically relevant dosage (for medication) or frequency (for therapy), we will conduct a sensitivity analysis for this aim using treatment (yes/no) as a time-varying covariate in the model to account for the potential effect of outside treatment. In this context, 'evidence-based' treatments will be operationally defined to include any psychosocial or medication treatments indicated for OCD or closely related anxiety disorders that have shown efficacy in prior clinical trials, and 'clinically relevant dosage' will be evaluated by psychiatrists or psychologists in our treatment program who specialize in OCD or related disorders. In addition, we will also add descriptive stats about how many participants are responders at both end-of-treatment and 12-month follow-up, how many participants are new responders by the 12-month follow-up, how many participants are no longer considered responders at the 12-month follow-up, and how many participants were not considered responders at neither assessment. No formal comparisons will be made between treatments in these numbers.

Exploratory endpoint 3: Treatment utilization for OCD concerns following the 3-month follow-up.

Analysis:

We will use logistic regression analysis to examine whether the likelihood of seeking further treatment for OCD during the treatment or follow-up phases (overall, combined) differs between participants who were assigned to the app-CBT and those assigned to the supportive attentional control condition. For this analysis, we will operationally define "treatment utilization" as any indication of participants seeking treatment for OCD during or following the treatment phase, regardless of whether the treatments sought are pharmacological or therapy-based, at sufficient or insufficient dosage/frequency, or considered evidence-based or not.

D. Power analysis

The sample size required to test a significant treatment response difference between the app-CBT and the control condition at the end of treatment is n=116 (n=58 per group). The sample size estimate is based on a single degree of freedom contrast in an ANOVA design, implemented in SAS for Windows version 9.4. The power model used a two-sided alpha=0.05, a power of 0.90, equal allocation of participants into both treatment arms, an assumption of an effect size of 0.76 comparing the two treatments at the end of treatment, as well as an assumption of a standard deviation of 1.1. The anticipated effect size of the app-CBT was conservatively based on the 95% lower confidence limit of the mean between-group effect size of CBT vs. psychological placebos reported in Ost et al. (g=1.29, 95% CI: [0.76–1.81]; Öst, Havnen, Hansen, & Kvale, 2015). We chose this conservative limit, because the effect sizes of recent, large RCTs of CBT or iCBT vs. various psychological control conditions have been quite variable, ranging from a high of d=1.12 in a comparison of iCBT against online non-directive supportive therapy (n=101; E.

Andersson et al., 2012) to d=0.55 in a trial of iCBT vs. internet-based progressive relaxation therapy (n=179; Kyrios et al., 2018) and d=0.52 in a trial of CBT vs. stress management training (n=73; Whittal, Woody, McLean, Rachman, & Robichaud, 2010). In the process of evaluating effect sizes from previous trials, we did not differentiate between internet-delivered CBT and face-to-face CBT, because a meta-analysis including 13 RCTs that directly compared these two treatment modalities in patients with various psychiatric and somatic conditions (total N=1053) showed a pooled effect size (Hedges' g) at post-treatment of -0.01 (95% CI: -0.13 to 0.12; G. Andersson, Cuijpers, Carlbring, Riper, & Hedman, 2014); of note, though, this direct comparison has not been done in trials with patients with OCD. Furthermore, in the wake of the COVID-19 pandemic, various negative mental health states and stressors – including anxiety, worry, depression, loneliness, financial strain, and trauma symptoms – have increased in the US population (e.g., Fowers & Wan, 2020). The variability of these negative mental health states and stressors will likely be higher than during non-pandemic times as multiple waves of the pandemic and state-wide measures of social interaction restrictions sweep the nation in the foreseeable future. In response, we used a standard deviation of 1.1 (instead of the standardized 1.0) in our power calculations to allow for a 10% increase in the variability of symptom severity estimates during the ongoing effects of the COVID-19 pandemic and its socioeconomic aftermath. Lastly, we assumed a conservative drop-out rate of 20% by end of treatment, based on the dropout numbers reported in a meta-analysis of 18 studies targeting self-help treatments for OCD, where studies with self-administered treatment without therapist contact had a drop-out rate of approximately 40%, compared to drop-out rates of 17-20% in treatments with minimal to low therapist contact (Pearcy, Anderson, Egan, & Rees, 2016); the meta-analysis by Ost et al. (2015) reported an estimated dropout rate of 15.5% from in-person CBT treatment.

VII. RISKS AND DISCOMFORTS

A. Complications of surgical and non-surgical procedures
Not applicable

B. Drug side effects and toxicities
Not applicable

C. Device complications/malfunctions
If the app, the supporting clinician/coach dashboard, or the HealthWatch website, malfunctions or does not work for a period of time, the patient may be unable to use the app/website to receive treatment or communicate with their coach. Similarly, if the clinician/coach dashboard or coach email were inaccessible, the coach would not be able to communicate with the patient. These risks could result in minor harms to users such as inconvenience or a delay in treatment. If device malfunctions resulted in the patient's in-app weekly questionnaire responses not being sent to the coach/clinician, this could result in a delay in clinical response to an elevated safety questionnaire. To mitigate risks of a temporary device malfunction, participants are given the study staff and investigators' contact information in the consent document. Thus, the clinical and study staff can communicate with the participant by phone call instead of through the app. As with in-person therapy, there are times when a therapy session gets cancelled and treatment, as well as weekly symptom assessments, are thus temporarily delayed. Likewise, in face-to-face therapy, a therapist may be temporarily out of reach at times (e.g., on vacation) and backup lines

of communication (e.g., pager coverage by a colleague) are put in place in case of emergencies. Thus, device malfunctions that lead to temporary delays in accessing the app functionality are similar to risks associated with temporary delays in traditional face-to-face therapy. See also, "Minimizing of Risks and Safety Reporting" under "D. Psychosocial (non-medical) risks" for further description of protections in place.

D. Psychosocial (non-medical) risks

Participants may feel uncomfortable due to the sensitive nature of the questions they may be asked. Likewise, some participants may feel uncomfortable about having assessment or coaching sessions digitally recorded and reviewed by project staff (which is necessary for rater supervision as well as assessment of the reliability of ratings adherence and competence). Participants could experience an increase in symptoms related to the natural waxing and waning of OCD symptoms. OCD can be associated with other psychiatric (e.g., depression, anxiety) symptoms, as well as suicidal thoughts and behaviors, which may also change over time. Breach of confidentiality, which great care will be taken to prevent, represents a potential risk. Participants may also feel uncomfortable about passive data collection via their smartphone. As discussed below, we will take precautions to ensure that these potential risks are minimized (see Adequacy of Protection Against Risks below).

Privacy and Confidentiality

All information gathered will be kept strictly confidential. We will adhere to the following procedures to protect privacy and confidentiality:

1) Significant effort will be invested in minimizing the risk of unauthorized access to study data and to mitigating the consequences of an unauthorized access were it to occur. To mitigate harm of unauthorized access and to increase confidentiality in the setting of authorized access, participants will be assigned a code number, which will be used instead of names or other identifiably information on paper forms and any email communication about CRFs between staff members. A link between ID number and participant's names will be kept in a separate secure password-protected file, saved on our secure MGH lab server and/or Partners Dropbox Business, as well as in the HIPAA-compliant REDCap study databases. Participants' names or other identifying information will not appear on any questionnaires, study documents, digital recordings, computerized data files, or published reports. Case records within REDCap that contain identifiable information will be reviewed only by study personnel or, if necessary, by institutional, sponsor-assigned, state, or federal regulatory personnel. Research assistants and others working on this study (e.g., technologists) will be educated about the importance of strictly protecting participants' rights to confidentiality.

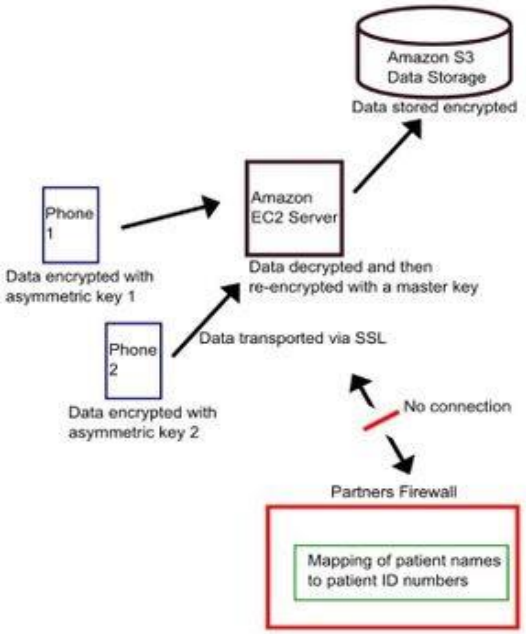
2) All personnel will be trained in research confidentiality procedures and HIPAA, including completion of CITI training and Healthstream training. Only the study personnel (or, if necessary, institutional, sponsor-assigned, state, or federal regulatory personnel) will have access to the identifiable data (see Figure 1, below).

3) Clinical data collected during MGH assessments: Computerized data (i.e., neurocognitive task data collected via Inquisit) will be stored in de-identified files on the protected lab servers.

Digital audio files of IE assessments will be stored in password protected files saved on the protected lab server. Paper research records (i.e., YBOCS and MINI assessment forms) will be kept de-identified and stored in locked file cabinets within locked offices in the MGH Center for OCD and Related Disorders. Electronic assessment forms in PDF format (i.e., optional method for collecting YBOCS and MINI assessment forms) will also be kept de-identified by using only participant code numbers instead of patient identifiable information. The PDF files will be saved in a study-specific Partners Healthcare Dropbox Business folder (for details about Dropbox Business data security, please see the next section "4) Technologist data"). Self-report measures and some clinician administered measures will be collected using REDCap. REDCap (Research Electronic Data Capture) is a free, secure, HIPAA compliant web-based application hosted by the Partners HealthCare Research Computing, Enterprise Research Infrastructure & Services (ERIS) group. Data collection projects rely on a study-specific data dictionary defined by members of the research team with planning assistance from Harvard Catalyst | The Harvard Clinical and Translational Science Center EDC Support Staff. The REDCap software allows researchers to design and implement study surveys for collecting, storing, retrieving, and manipulating data electronically. Once built, participants and authorized study staff identified in the delegation of authority log can enter data directly into REDCap surveys via any computer or tablet with standard web access and browsers. Any data that is transmitted electronically to the REDCap server is encrypted. Participants entering survey data through a web-browser will only have access to their own current survey, but not their past survey data or data entered by any other participant. Data fields containing names, contact information, social security numbers, or date of birth are identified as "Identifiers" in the database, which can easily be removed during data exports to create 'de-identified' datasets in compliance with the Standards for Privacy of Individually Identifiable Health Information ("Privacy Rule") of the Health Insurance Portability Act of 1996 (HIPAA). Data within Partners is stored automatically and securely on an MS SQL Server, accessed over industry standard SSL 128 bit RSA encryption during data transfers. Data is routinely backed up locally onto a redundancy server and stored in a separate database that is locked with 256 AES encryption. Long term storage on Partners servers occurs nightly and allows for incremental backup over multiple systems. Therefore, should one drive be physically damaged, there will be multiples within the chain to replace it. Both data servers are stored within PHS IS corporate firewall, in a secure, key access facility with password protected computers. Only vetted PHS security officials will have access to physical machines storing study data. Since data are stored on a protected server, a compromise of any individual computer at a research facility will not lead to a breach of the secure database.

4) Technologist data: Computerized data and digital audio files collected from feedback interviews with technologists will be stored de-identified, in password protected files in the Partners Healthcare Dropbox Business folder. The PHS Dropbox Business folder will be set up by a listed owner with a Partners email address. Per Partners Research Information Services and Computing, "The enterprise rollout of Dropbox Business at Partners HealthCare is an approved storage and collaboration solution. This version of Dropbox Business provides unlimited storage, fully encrypted data (AES-256 encrypted) and is compliant with Partners' policies and procedures. Dropbox Business allows you to sync, share, and manage your files online" (https://rc.partners.org/kb/article/2285).

5a) Data collected through the app: All intervention app data will be sent from the participant's mobile client device to the clinical study server via an Internet connection secured by SSL. The clinical study server will be located in the PHS/MGH Secure Data Centers and hosted on PHS managed infrastructure. The collected data will be linked to the study identifier only (described in "1" above). Participants will be advised to avoid identifying themselves in answers to free response questions. The administrative interface to the clinical study server (channels 1 and 5 below) will be password protected with access limited to study staff. To further minimize risk the administrative interface of the clinical study server will only be available over the PHS network.



5b) Data collected through the control site: Interactions between participant's client device and the control condition site will occur over the Internet using secured by SSL to secure the data in transit. The control site will be hosted on the MGH OCD Programs commercial WordPress account. The only data persisted in the control condition site will be the anonymous study identifier, the survey week identifier, and confirmation of survey completion. Although surveys are presented for participants to complete the actual answers are not part of the evaluation and thus they are not saved for analysis in the control condition to minimize inadvertent disclosure risk. The control site will be password protected to limit access to study staff.

6) Beiwe: The Beiwe study app and platform will be used to collect passive smartphone data. Beiwe is a secure, HIPAA-compliant research platform for digital phenotyping developed by Dr. Onnela (collaborator, Harvard School of Public Health). Beiwe includes (a) a backend to collect and store raw passive data from participants' personal Android and iOS smartphones; (b) a participant-facing app adapted for individual studies, and (c) a data analysis pipeline that pre-processes raw mobile data into validated summary statistics. Each study has a unique app that is separate from all other studies using Beiwe, and which is managed by that study's researchers. The unique study app is configured to collect the passive and active data specified for that study's needs, and nothing more (e.g., particular data streams, sampling frequency). To encourage occasional interaction with the Beiwe app, we will administer a weekly survey question to participants via Beiwe (Wifi check-in question).BEIWE data are stored in a restricted-access, HIPAA-compliant Amazon Web Service (AWS) server operated by Dr. Onnela's lab at the Harvard School of Public Health. Key features are summarized below:

- Participants are enrolled into their unique Beiwe study with a de-identified 8-character user ID (e.g., yixg8437) and temporary password, that allows them to download the app specific to their study.
- Participants can only download and use the study's Beiwe app if given the proper registration code from the study investigator (e.g., once fully consented and determined to be eligible).

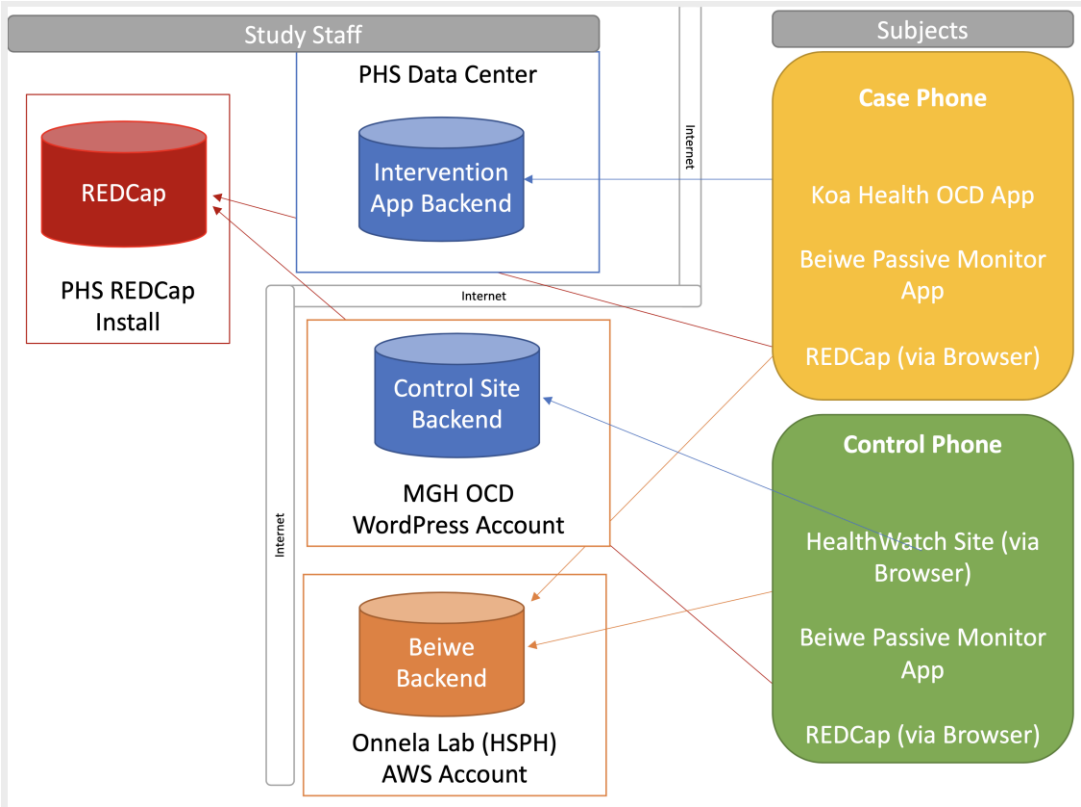
- Registration codes can only be used on a single phone.
- To use the app, participants must log in with their de-identified user ID and a password, which must be at least 6 characters long. All data collection features of the app are protected behind this login wall. The app logs out automatically if there is no activity for a configurable number of minutes (e.g., 5 minutes).
- All data are stored with the de-identified user ID. No identifying data such as name or contact information are tied to participants' data in Beiwe or stored in the Beiwe app.
- Indirect identifying data (e.g., telephone numbers from phone and text logs) are hashed using an industry-recognized strong hashing algorithm (SHA-256), to render all data unidentifiable.
- GPS data will be collected using a "fuzzing" procedure that adds random latitude and longitude offsets to GPS coordinates. In this way, actual latitude and longitude coordinates are not collected.
- All data are encrypted using industry-standard encryption techniques, both in transit and at rest. Data are not stored on participants' phones unencrypted. The phone also uses asymmetric encryption, meaning that even the phone cannot read its own data; data recorded on the phone can only be read on the server.vi
- During registration the device is provided with the public half of a 2048 bit RSA encryption key. With this key the device can encrypt data, but only the server, which has the private key, can decrypt it. The RSA key is then used to encrypt a symmetric AES key for bulk encryption. These keys are generated as needed by the app, are not stored, and must be decrypted by the server before any data can be recovered. Data received by the server are then re-encrypted with a master key provided for that study, and then stored on Amazon S3, an industry-standard secure storage platform housed in data centers that are protected by armed guards.
- Amazon Web Services (AWS) has released a whitepaper (attached) describing how EC2 and S3, the two Amazon services Beiwe uses, meet HIPAA compliance standards. Access to the AWS account on which data are stored is restricted and requires login credentials.
- All data connections to the web service hosting the study are negotiated on industry-standard SSL/TLS connections, removing the vulnerability of man-in-the-middle attacks or packet-sniffing data leaks.
- Below is a visual of the data encryption system including the phones, Amazon servers, and the separation of participant information behind a Partners Firewall.

The procedures we will follow to protect participants' data security and confidentiality are consistent with those used in previous research studies at our institution and have been highly effective in protecting participants' confidentiality and data security

6) Telehealth assessments will be conducted using the Partners-approved platform for clinical video calls, which has been vetted by Partners to be secure and HIPAA-compliant. Participants will be instructed in advance to be in a private location (e.g., in a room with closed doors) during the calls.

7) Only de-identified data will be shared with the sponsor (Koa Health), under the terms of the Data Use Agreement (i.e., Statement of Work).

Figure 1. Data flow and storage chart –Overview of All Data Sources.



See Figures 2-14 in "Independent Monitoring of Source Data" for a detailed diagram and description of the data flow and storage described above in Figure 1. The administrative interface to this system will only be accessible to work stations on the PHS/MGH network. Of note, the CBT app architecture and code base were previously evaluated in #2017P000293, which is exactly equivalent to that described here. Similarly, the substantially equivalent REDCap and Beiwe flow and storage have been previously evaluated in #2017P000293 and #2019P002041 respectively.

Minimizing of Risks and Safety Reporting

The following procedures will be implemented to protect participants against risks. The information provided in this section pertains to all study phases, unless otherwise indicated.

1. Participants with active suicidal ideation at the screening assessment will be excluded from participating (see Inclusion/Exclusion Criteria). If a subject scores ≥21 on the

QIDS-SR and/or >1 on the self-report suicide questionnaire, the independent evaluator will follow up with a phone evaluation within 24 hours and refer the participant to a higher level of care if clinically indicated. Suicidal ideation at screening is also measured using the C-SSRS (see Inclusion/Exclusion criteria), and if a potential participant scores ≥2, a study clinician, PI, or independent evaluator will conduct a risk evaluation with the participant and refer the participant to a higher level of care if clinically indicated.

2. Participants with severe major depressive disorder will be excluded from participating (see Inclusion/Exclusion criteria).

3. A disclaimer that is accessible from the home page of the digital CBT program and control condition website will be presented to remind participants that if they are experiencing suicidal thoughts, they should seek professional help or go to the emergency room right away. Links to 988, 911 and suicide hotline numbers will be provided along with this disclaimer.

4. A general resources page will be available on the app and website at all times to participants, which will include a suicide hotline number.

5. Participants' clinical improvement or deterioration will be assessed weekly via a participant-rated CGI-OCD collected via the app in the app-based CBT group or REDCap surveys in the supportive attentional control condition. Participants may be withdrawn from the study if their clinical condition deteriorates substantially. Deterioration will be defined by a rating of 6 (much worse) or 7 (very much worse) on the patient –rated CGI during three consecutive, weekly assessments and (2) PI judgment that remaining in the study is not in the participant's best interest. Of note, three consecutive weekly ratings of 6 or 7 on the weekly, participant-rated CGI-OCD will also trigger a notification to the clinician via text message or email. In the case that a BA-level coach is notified that a participant's CGI indicates deterioration, they will notify a doctoral-level clinician as soon as possible (and within 24 hours). A study clinician will follow up with a phone evaluation within 24 hours of the alert and refer the participant to a higher level of care if clinically indicated.

6. Ratings on a one-item self-report suicide questionnaire will be carefully monitored weekly via the app in the app-based CBT group or REDCap surveys in the supportive attentional control condition; a score >0 at any assessment will trigger a pop up message to be presented to the patient within the mobile app reminding participants that if they are experiencing suicidal thoughts, they should seek professional help or go to the emergency room right away. Links to 988, 911, as well as a national suicide hotline, will be provided within this popup notification. A score >1 will <u>also</u> trigger notification to the clinician/coach via text message or email. In the case that a BA-level coach is notified that a participant reported a score>1, they will notify a doctoral-level study clinician as soon as possible (and within 24 hours). A study clinician, PI, or independent evaluator will follow up with a phone evaluation within 24 hours of the alert and refer the participant to a higher level of care if clinically indicated. To be able to determine which participant triggered the alert (i.e., to link the de-identified trigger with the actual

participants' name) and to be able to contact the person (i.e., to be able to look up that person's telephone number), the clinicians on call will require remote access to the key that links the de-identified study ID with the patient identifier. We will use Partners Dropbox Business to store this password-protected Excel spreadsheet that links the participants' de-identified and identified data. Partners Dropbox Business is a secure and appropriate platform for storing participant de-identified and identified data (see attached RISO approval). Access to this secure spreadsheet will only be provided to those who are directly responsible for risk assessment and will be provisioned/deprovisioned accordingly as clinicians join or depart the study.

7. Participants may also be withdrawn if, in the judgment of the PI, remaining in the study poses a substantial risk to the participant or a higher level of care is needed.

8. Treatment through app-based CBT or HealthWatch will be supplemented with communication with a study coach, who can answer questions and guide participants through the treatment as needed. Study coaches will be trained BA-level research assistants who will supervised weekly by licensed, doctoral level clinicians who are experts in OCD and related disorders. See *In-App Coach Training and Qualifications*, above, for details.

9. The independent evaluator(s) will be highly experienced, highly trained, and closely supervised.

10. Dr. Wilhelm will be available, if necessary, to discuss the study, alternative treatments, or any concerns about the study with participants if requested by the participant, coach or rater.

11. Drs. Wilhelm, Weingarden, Greenberg, and the participants' study clinician/coach will be available to answer study questions via the app or phone. This will be clearly communicated orally and in writing to study participants.

12. All participants will be provided with referral resources.

13. The study clinicians. coaches, and raters will make every attempt to help participants feel comfortable when discussing sensitive material. Participants may skip questions on assessments that they are uncomfortable answering.

14. If exposure exercises suggested through the CBT app are too anxiety provoking, participants will be able to do alternative exercises that cause less anxiety.

15. The CBT treatment will initially emphasize cognitive restructuring, which we anticipate will be less anxiety provoking than exposure treatment alone and will make exposure more tolerable.

16. Technologists from Koa Health who conduct the technology feedback interviews are highly trained professional staff with experience conducting patient interviews and user-

centered design. They have received additional training (including a multi-day workshop) from our team (The Center for OCD and Related Disorders) on CBT, and issues of confidentiality. They have completed CITI training.

17. Three clinical psychologists or researchers familiar with OCD will be selected to serve as a Data Safety Monitoring Board, to review the study once a year.

18. The subject will designate a relative or friend who could be contacted should the subject be unavailable and the investigator has concerns about the subject's well-being.

19. The digital application will be scanned by Veracode prior to deploying the app to participants.

20. Participants will not be asked to refrain from seeking treatment between the 3-month follow-up assessment and the 12-month follow-up assessment.

We anticipate that the above procedures will be effective in protecting study participants against potential risks.

Adverse event reporting:
See below: "Adverse event reporting guidelines"

E. Radiation risks
Not applicable

VIII. POTENTIAL BENEFITS

A. Potential benefits to participating individuals
Participants may benefit from the comprehensive diagnostic assessment with a clinician assessor. Participants may benefit from careful clinical monitoring and potentially effective treatment. This study has the potential benefit of improving the patients' OCD symptoms.

B. Potential benefits to society
If app-based CBT for OCD is effective, it may offer increased, cost-effective access to CBT for OCD, a treatment that is empirically supported and is otherwise difficult to access.

IX. MONITORING AND QUALITY ASSURANCE

A. Independent monitoring of source data. Grace Bartoo and employees of Decus Biomedical, Inc. were hired by Koa Health to serve as study monitors. They may conduct visits up to 4 times yearly and will review study data to provide feedback regarding ongoing study procedures as they relate to any potential future pursuit of FDA clearance for the *Perspectives* app.

B. Study staff monitoring of source data
The PI will have overall responsibility for study data and participant safety. Please see "Privacy and Confidentiality" above for more information about the data collected for the present study.

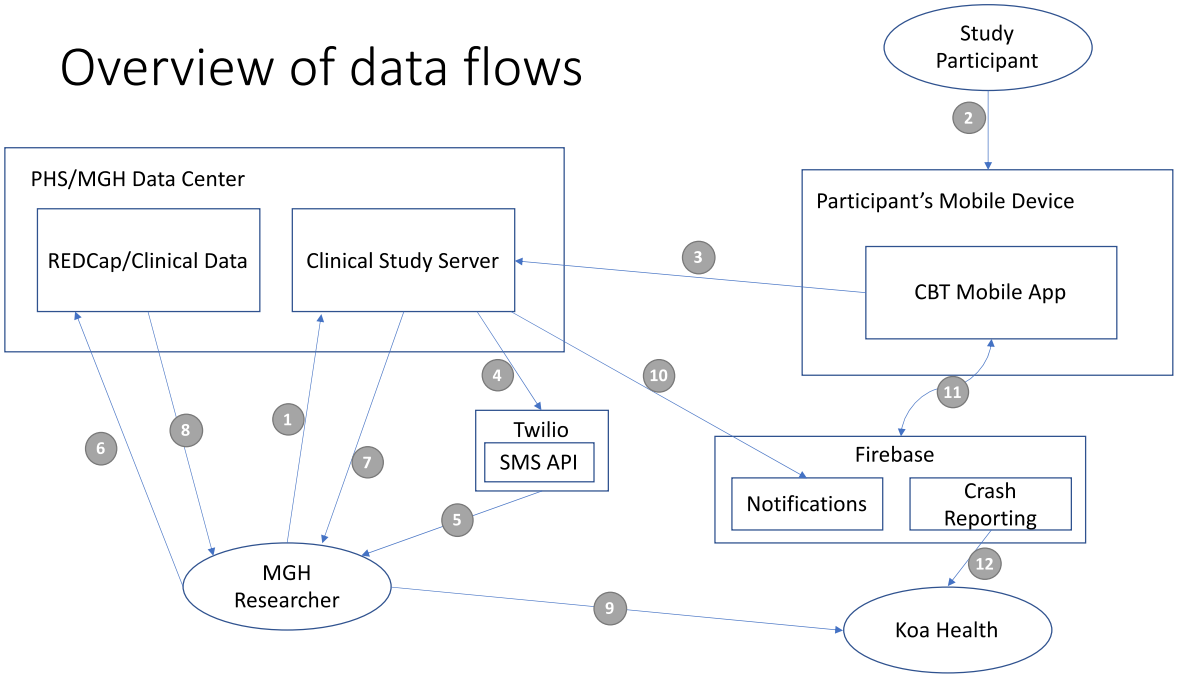
All aspects of the study will be conducted in accordance with the hospital's policy on confidentiality and applicable ICH Good Clinical Practice (GCP) guidelines to ensure protocol adherence, quality of data, study treatment accountability, and compliance with regulatory requirements.

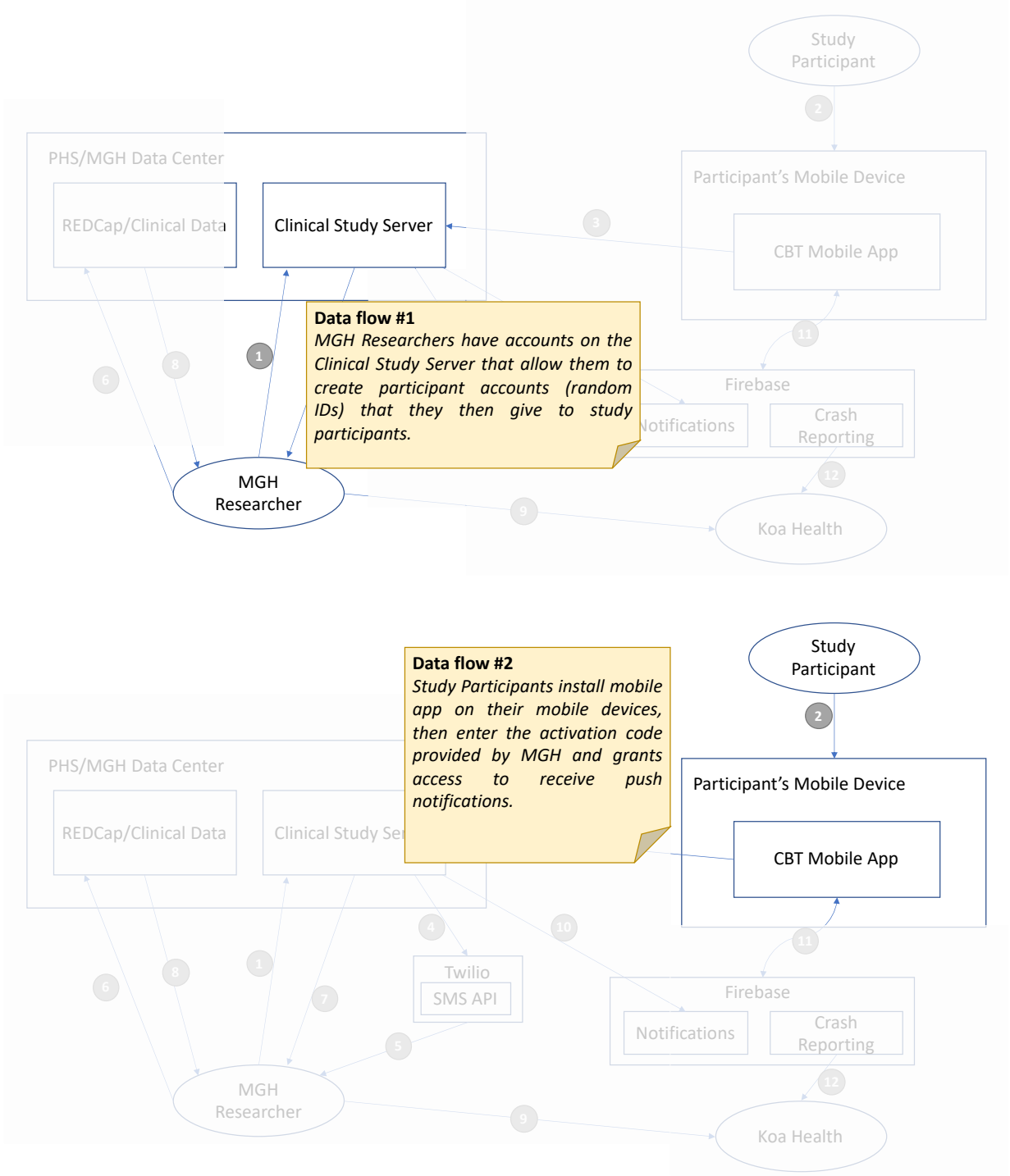
The majority of study assessments (see "Measures Chart OCD" on pages 18-19) will be completed in REDCap. REDCap provides flexible features that can be used for a variety of research projects and provides an intuitive interface to enter data with real time validation (automated data type and range checks). The system offers easy data manipulation with audit trails, reports for monitoring and querying participant records, and an automated export mechanism to common statistical packages (SPSS, SAS, Stata, R/S-Plus). We will use an iterative development and testing process to test our data collection strategy and implementation for all assessment forms used in this study, including form-specific automatic data validation checks (e.g., date vs. integer validation), data range checks (e.g., minimum and maximum plausible dates), missing item on-screen notifications, and forced-choice fields (e.g., yes/no or multiple choice answers). All study staff using REDCap will have defined roles and privileges in line with their delegated tasks as pre-determined by the database manager in consultation with the PI and Sub-Investigators. IEs will enter their clinical assessment data directly into REDCap, except for two forms (i.e., YBOCS and MINI) that will first be collected as paper forms or as electronic records in PDF format. The electronic forms of the MINI and YBOCS assessments will be completed by IEs using the Notability software on iPads, where the assessment data is entered using a stylus and the PDF record plus data is saved as a new PDF file. The completed PDF-copies of the MINI and YBOCS forms are then uploaded into the study-specific Partners Healthcare Dropbox Business folder designated for this purpose. One study staff member (either the IE or a designated, trained RA) will then enter the data from the paper or PDF forms into the designated REDCap forms within 3-5 business days of the completion of the assessment visit, and another staff member (trained RA) will enter the same forms into double data entry records for data entry comparisons. If discrepancies are noted, the RA responsible for data entry will check the source documentation and correct the REDCap form as necessary. Data entry by participants for self-report measures in this electronic data capture system precludes the need for subsequent data entry by staff, thus minimizing human error, and resulting in improved data integrity and quality. Patients will enter survey responses into an electronic assessment form on subject-facing REDCap, and the responses will then be transmitted and stored in the secure REDCap database. Together with the data collected on paper forms or PDF records (i.e., MINI and YBOCS), via the Inquisit website, the *Perspectives* app, Beiwe, and the HealthWatch website (see sections "Privacy and Confidentiality" above for details), the data entered by IEs and RAs into REDCap will be considered the source documentation.

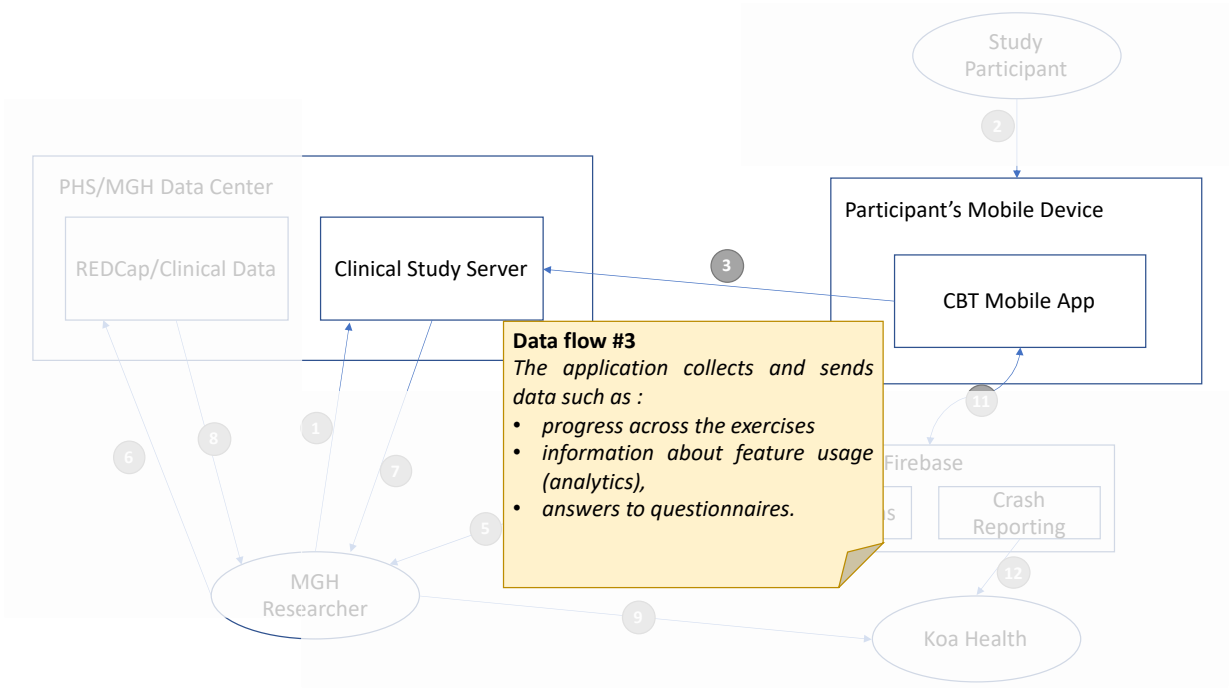
The RA responsible for data entry will inspect each form completed by the IE and the study participant within 1-3 business days of the assessment; if required data elements in IE completed forms are missing, the RA will contact the IE to complete the missing items. If participant-completed surveys are missing key variables (i.e., basic demographics, ratings used to evaluate safety based on suicidal ideation or symptom worsening), the RA will contact the participant by email or phone to obtain the missing data.

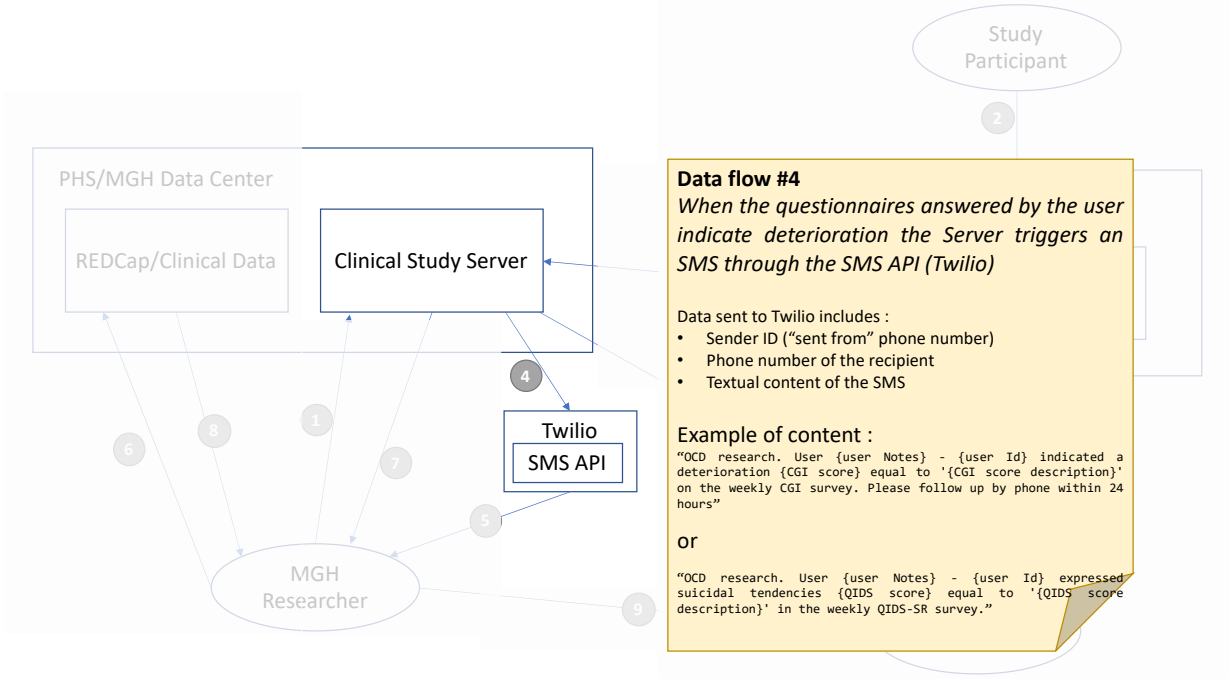
The data manager or an RA trained in data management tasks will run reports of the available REDCap data summarizing aggregate accrual and retention information, adverse events, protocol deviations, and inter-rater reliability ratings to track protocol adherence and check data quality with the RA responsible for data entry. These reports will be run on a weekly or biweekly basis. The PI and Sub-Investigators will review these reports once a month. In addition, the data manager will run monthly systematic data checks of the REDCap databases to check for inconsistent data values, missing data fields, missing forms, and study visits conducted out of window. Any failed data checks will be brought to the attention of the RA responsible for data entry, who will resolve them in REDCap within 1-14 business days. Changes made to the REDCap source documentation will be tracked within REDCap through data history logs that keep timestamps, user-IDs, as well as past and new data entered in an automated audit trail.

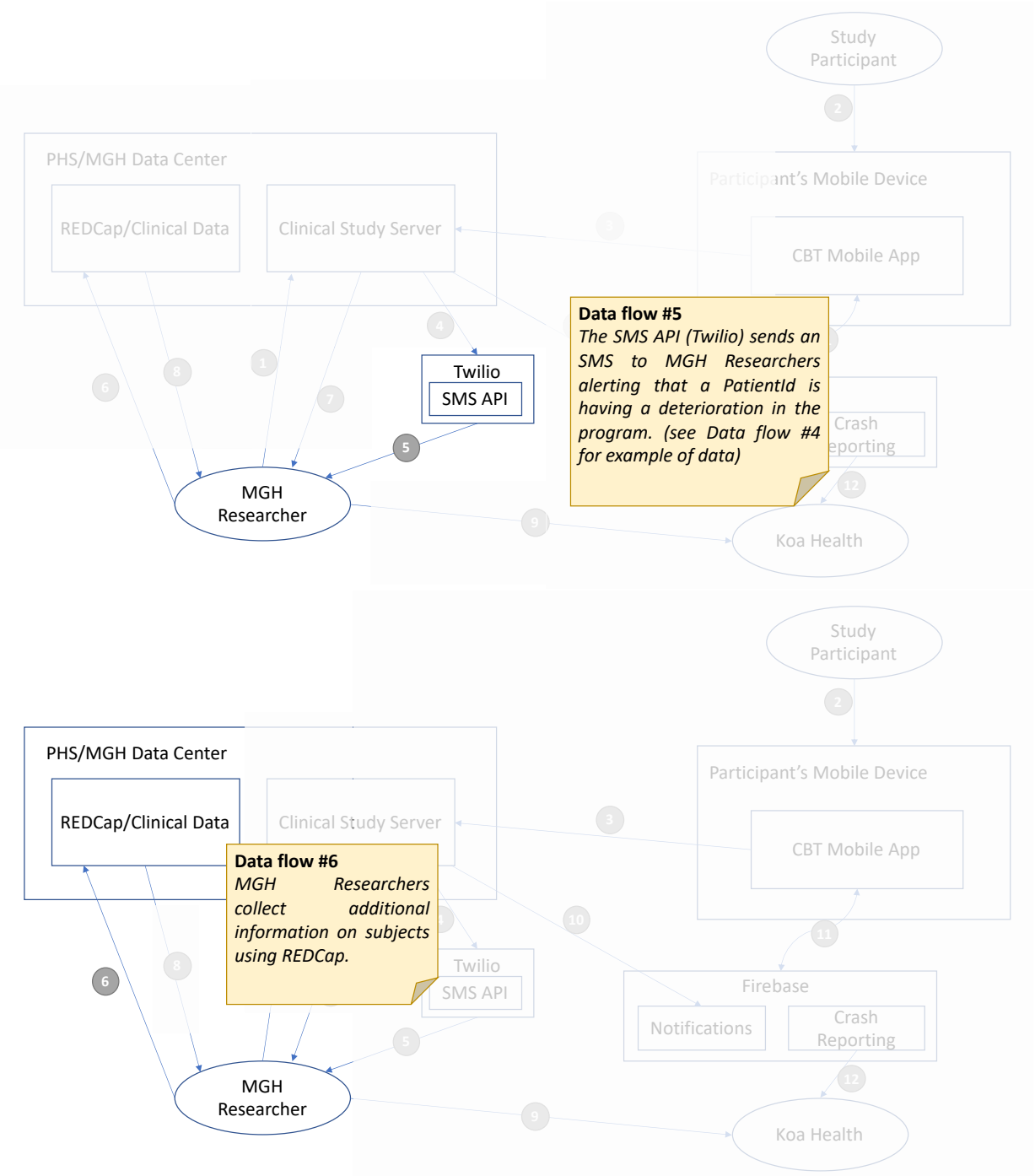
Figures 2-14-. Data flow and storage chart – Overview and Breakdown of App Data

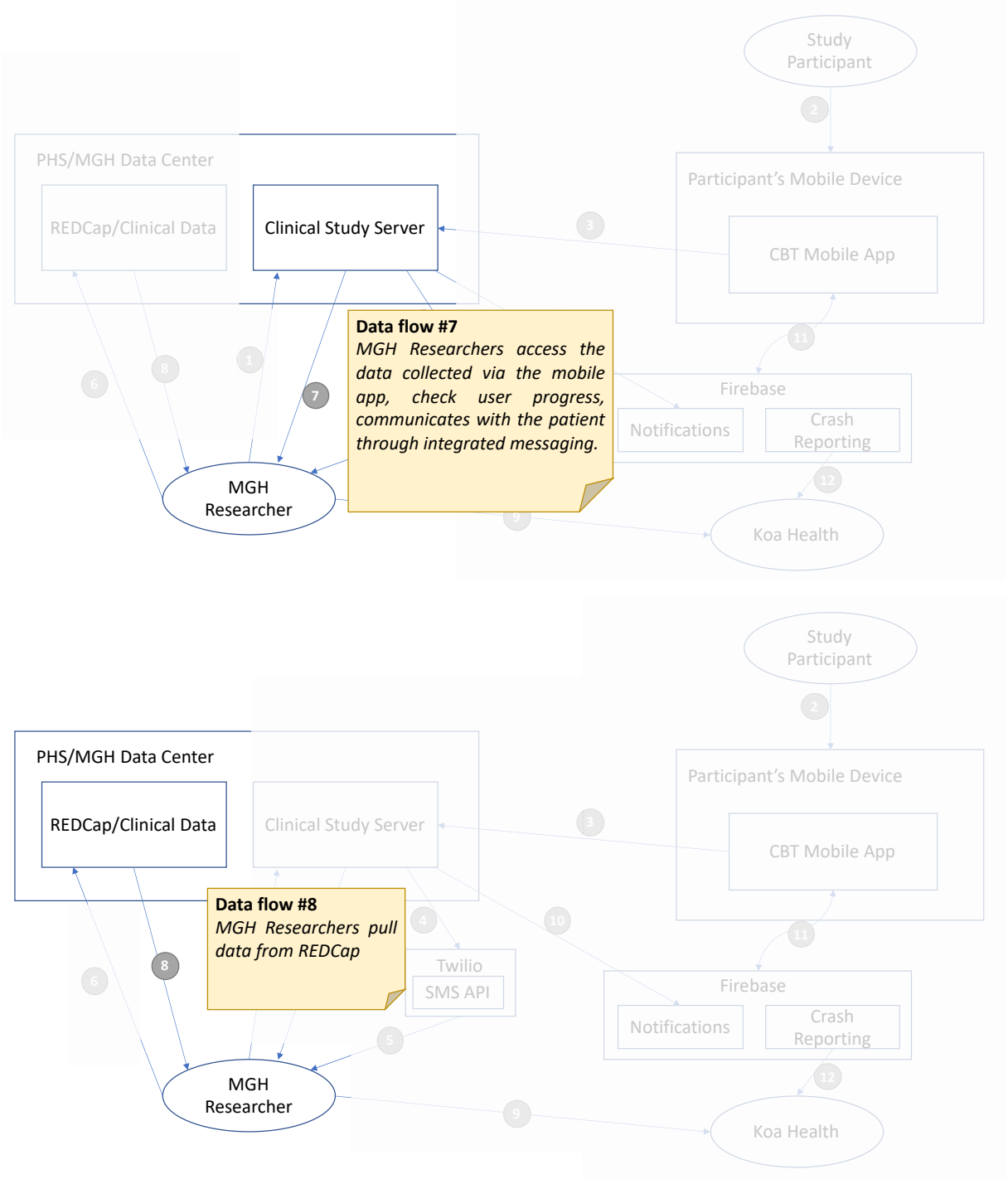


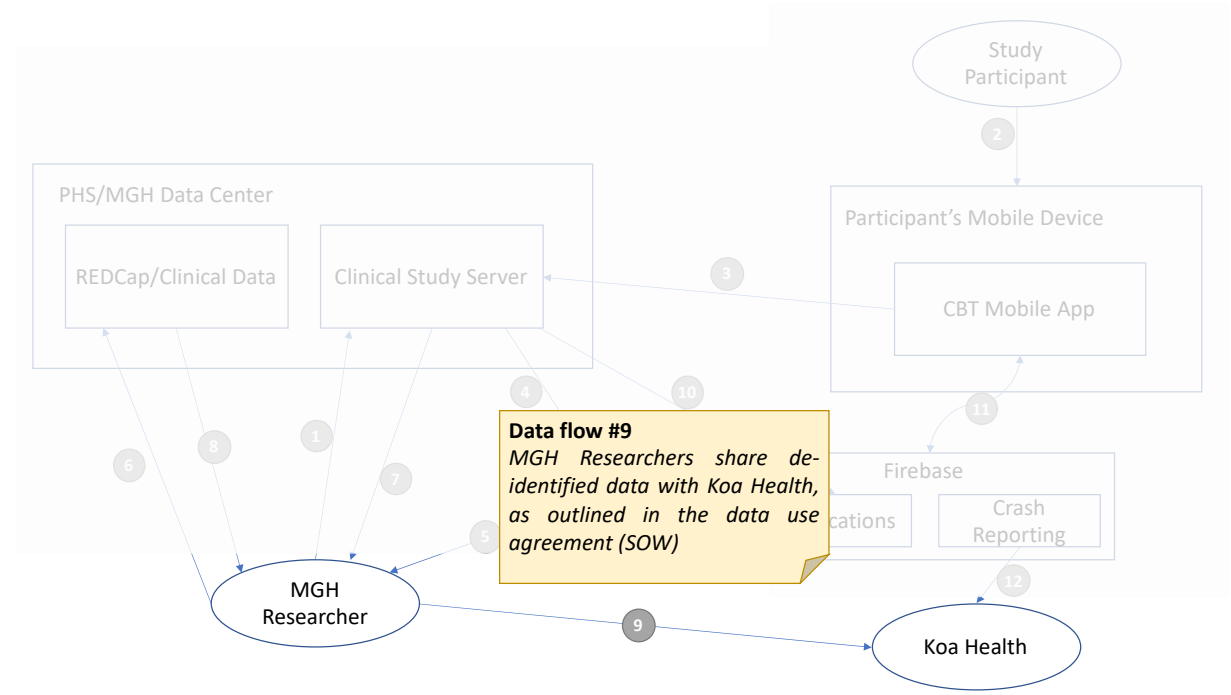

Study Participant
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Participant's Mobile Device
CBT Mobile App
Firebase
Notifications
Crash Reporting
MGH Researcher
Koa Health
Twilio
SMS API
Data flow #1
MGH Researchers have accounts on the Clinical Study Server that allow them to create participant accounts (random IDs) that they then give to study participants.
Data flow #2
Study Participants install mobile app on their mobile devices, then enter the activation code provided by MGH and grants access to receive push notifications.

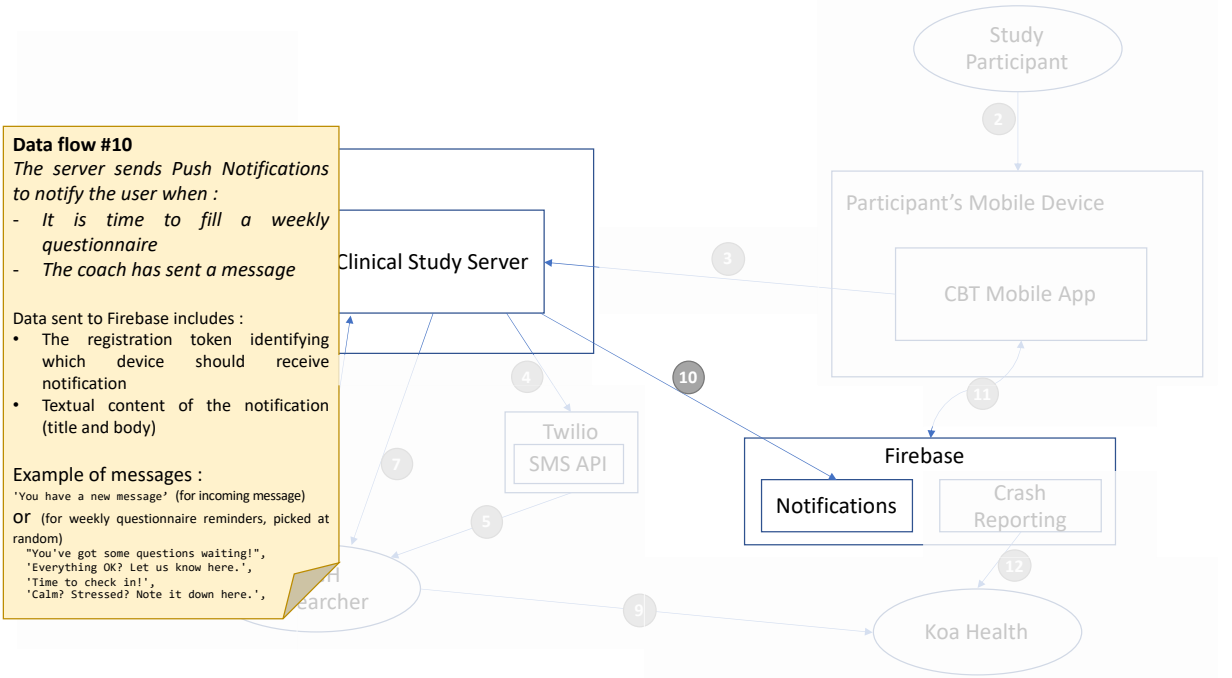

Study Participant
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Participant's Mobile Device
CBT Mobile App
Firebase
Crash Reporting
MGH Researcher
Koa Health
Data flow #3
The application collects and sends data such as :
• progress across the exercises
• information about feature usage (analytics),
• answers to questionnaires.

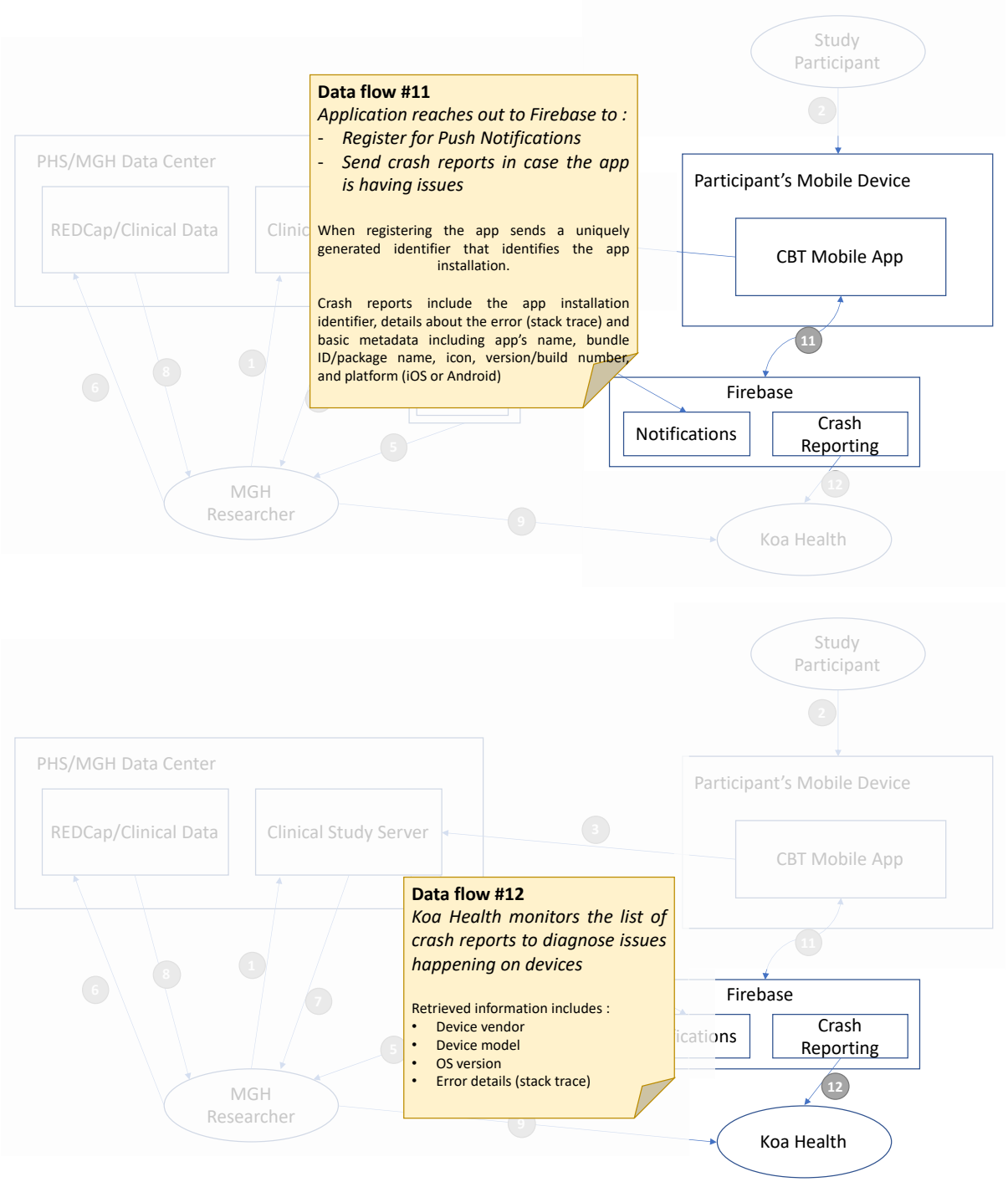

Study Participant
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Twilio
SMS API
MGH Researcher
Data flow #4
When the questionnaires answered by the user indicate deterioration the Server triggers an SMS through the SMS API (Twilio)
Data sent to Twilio includes :
• Sender ID ("sent from" phone number)
• Phone number of the recipient
• Textual content of the SMS
Example of content :
"OCD research. User {user Notes} - {user Id} indicated a deterioration {CGI score} equal to '{CGI score description}' on the weekly CGI survey. Please follow up by phone within 24 hours"
or
"OCD research. User {user Notes} - {user Id} expressed suicidal tendencies {QIDS score} equal to '{QIDS score description}' in the weekly QIDS-SR survey."


Study Participant
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Participant's Mobile Device
CBT Mobile App
Twilio
SMS API
MGH Researcher
Crash Reporting
Koa Health
Data flow #5
The SMS API (Twilio) sends an SMS to MGH Researchers alerting that a PatientId is having a deterioration in the program. (see Data flow #4 for example of data)
Study Participant
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Participant's Mobile Device
CBT Mobile App
Data flow #6
MGH Researchers collect additional information on subjects using REDCap.
Twilio
SMS API
Firebase
Notifications
Crash Reporting
MGH Researcher
Koa Health


Study Participant
2
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Participant's Mobile Device
CBT Mobile App
3
11
Data flow #7
MGH Researchers access the data collected via the mobile app, check user progress, communicates with the patient through integrated messaging.
Firebase
Notifications
Crash Reporting
12
6
8
1
7
MGH Researcher
9
Koa Health
Study Participant
2
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Participant's Mobile Device
CBT Mobile App
3
11
Data flow #8
MGH Researchers pull data from REDCap
4
10
Twilio
SMS API
Firebase
Notifications
Crash Reporting
12
6
8
5
MGH Researcher
9
Koa Health


Study Participant
PHS/MGH Data Center
REDCap/Clinical Data
Clinical Study Server
Participant's Mobile Device
CBT Mobile App
Firebase
Crash Reporting
MGH Researcher
Koa Health
Data flow #9
MGH Researchers share de-identified data with Koa Health, as outlined in the data use agreement (SOW)


Study Participant
Participant's Mobile Device
CBT Mobile App
Clinical Study Server
Twilio
SMS API
Firebase
Notifications
Crash Reporting
Koa Health
Data flow #10
The server sends Push Notifications to notify the user when :
- It is time to fill a weekly questionnaire
- The coach has sent a message
Data sent to Firebase includes :
• The registration token identifying which device should receive notification
• Textual content of the notification (title and body)
Example of messages :
'You have a new message' (for incoming message)
or (for weekly questionnaire reminders, picked at random)
"You've got some questions waiting!",
'Everything OK? Let us know here.',
'Time to check in!',
'Calm? Stressed? Note it down here.',



De-identified data collected by the study's Beiwe app will be stored in a guarded, restricted-access Amazon Web Service (AWS) account operated by a collaborator at the Harvard School of Public Health. AWS has released a whitepaper (attached) describing how EC2 and S3, the two Amazon services Beiwe uses, meet HIPAA compliance standards. Access to the AWS account on which data are stored is restricted and requires username and password. Participants may withdraw their data by notifying the PI in writing.

We anticipate that the above procedures will ensure the confidentiality and integrity of study data.

C. **Ongoing Monitoring of Coaching Fidelity**
All coach-participant telephone interactions will be digitally audio recorded. Once coaches meet the initial certification standard (see above), a doctoral level independent adherence rater will rate all sessions at regular intervals during the study using adherence and competence measures adapted from prior trials. This will be done to ensure adherence to the coaching manuals, competent delivery of coaching, and to prevent cross-contamination (e.g., coaches providing CBT-specific guidance to participants in the HealthWatch condition). The adherence rater will be trained and supervised by Drs. Wilhelm, Greenberg, or Weingarden. Descriptive statistics on adherence and competence ratings will be obtained.

If minimum standards are not met (i.e., if two consecutive recorded sessions receive an adherence or competence rating below the above certification standard), the coach will receive additional training from Drs. Wilhelm, Greenberg, or Weingarden, and the next two consecutive sessions will be reviewed and must meet certification standards for continued coaching to occur. Coaches who do not meet these standards will be replaced (although we do not expect this to occur, given that coaches will be extensively trained and we will provide ongoing supervision).

D. Safety monitoring
Three clinical psychologists or researchers knowledgeable about OCD will be selected to serve as a Data Safety Monitoring Board, to review the study once a year when the study is actively enrolling. The PI will have overall responsibility for monitoring the integrity of study data and participant safety. Procedures for managing participant safety, including the monitoring of participants throughout the trial and response to clinical deterioration (as defined above) should it occur, are detailed above in "**Minimizing of Risks and Safety Reporting.**"

E. Outcomes monitoring
Adverse events and data completeness will be monitored regularly throughout the study as described above.

F. Adverse event reporting guidelines

Adverse event reporting:
Adverse events will be reported per PHRC guidelines.

X. REFERENCES

Abramowitz, J. S. (1996). Variants of exposure and response prevention in the treatment of obsessive-compulsive disorder: A meta-analysis. Behavior Therapy, 27(4), 583-600. DOI: 10.1016/S0005-7894(96)80045-1

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Andersson, E., Enander, J., Andrén, P., Hedman, E., Ljótsson, B., Hursti, T., Bergström, J., Kaldo, V., Lindefors, N., Andersson, G., Rück, C. (2012). Internet-based cognitive behaviour therapy for obsessive–compulsive disorder: A randomized controlled trial. *Psychological Medicine, 42*(10), 2193-2203. doi:10.1017/S0033291712000244

Bang, H., Ni, L., & Davis, C. E. (2004). Assessment of blinding in clinical trials. *Controlled clinical trials*, *25*(2), 143-156.

Ben-David, B. M., Nguyen, L. L., & van Lieshout, P. H. (2011). Stroop effects in persons with traumatic brain injury: selective attention, speed of processing, or color-naming? A meta-analysis. *J Int Neuropsychol Soc, 17(02)*, 354-363.

Borkovec, T. D,. & Nau, S. D. (1972). Credibility for analogue therapy rationales. J Behav Ther Exp Psychiatry, 3, 257-260.

DeRubeis, R.J., Webb, C.A., Tang, T.Z., & Beck, A.T. Cognitive therapy. In: Dobson KS, ed. Handbook of cognitive-behavioral therapy. 3rd ed. New York: Guilford, 2010.

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry, 31,* 47-55.

Eddy, K. T., Dutra, L., Bradley, R., & Westen, D. (2004). A multidimensional meta-analysis of psychotherapy and pharmacotherapy for obsessive-compulsive disorder. Clinical Psychology Review, 24(8), 1011-1030.

Endicott, J., Nee, J., Harrison, W., & Blumenthal, R. (1993). Quality of Life Enjoyment and Satisfaction Questionnaire: A new measure. *Psychopharmacology Bulletin, 29*, 321-326.

Goodman WK, Price LH, Rasmussen SA, et al. The Yale-Brown Obsessive Compulsive ScaleI. Development, Use, and Reliability. Arch Gen Psychiatry. 1989;46(11):1006–1011.

Grant, D.A., & Berg, E.A. (1981) Wisconsin Card Sorting Test. Odessa, Fl: Psychological Assessment Resources, Inc.

Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, et al. (2012) The Effectiveness of an Online Support Group for Members of the Community with Depression: A Randomised Controlled Trial. *PLOS ONE* 7(12): e53244.

Guy, W. (1976). Clinical Global Impressions. In W. Guy (Ed.), *ECDEU Assessment Manual for Psychopharmacology (*rev. ed.; DHEW publication [ADM] 76-338; pp. 217-222*)*. Rockville, MD: U.S. Department of Health, Education, and Welfare, National Institute of Mental Health.

Kroenke, K. & Spitzer, R.L. (2002). The PHQ-9: A new depression and diagnostic severity measure. *Psychiatric Annals, 32*, 509-521.

Lowe B, Decker O, Muller S, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Medical care. Mar 2008;46(3):266-274. PMID: 18388841

Mahoney, Mackenzie, Williams, Smith, & Andrews. (2014). Internet cognitive behavioural treatment for obsessive compulsive disorder: A randomised controlled trial. Behaviour Research and Therapy, 63, 99-106.

McClintock, S. M., Husain, M. M., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Trivedi, M. H., … Rush, A. J. (2011). Residual symptoms in depressed outpatients who respond by 50% but do not remit to antidepressant medication. *Journal of clinical psychopharmacology*, *31*(2), 180–186. doi:10.1097/JCP.0b013e31820ebd2c

McConnaughy, E. A., Prochaska, J. O., & Velicer, W. F. (1983). Stages of change in psychotherapy: Measurement and sample profiles. *Psychotherapy: Theory, Research & Practice*, *20*(3), 368.

McMurtry, S. L. & Hudson, W. W. (2000). The Client Satisfaction Inventory: Results of an initial validation study. *Research on Social Work Practice, 10*, 644-663.

Mojtabai R (2005). Trends in contacts with mental health professionals and cost barriers to mental health care among adults with significant psychological distress in the United States : 1997–2002. American Journal of Public Health 95, 2009–2014.

Mundt, J., Marks, I., Shear, M., & Greist, J. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. British Journal Of Psychiatry, 180, 461-464.

Posner, K., Brent, D., Lucas, C., Gould, M., Stanley, B., Brown, G., ... & Mann, J. (2008). Columbia-suicide severity rating scale (C-SSRS). New York, NY: Columbia University Medical Center.

Rush, A. J., Trivedi, M. H., Ibrahim, H. I., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S. G., Manber, R., Thase, M. E., Kocsis, J. H. & Keller, M. B. (2003a).

The 16-item Quick Inventory of Depressive Symptomatology (QIDS) Clinician Rating (QIDS-C) and Self-Report (QIDS-SR): Psychometric evaluation in patients with chronic major depression. *Biological Psychiatry,* 54, 573–583.

Sareen J, Jagdeo A, Cox BJ, Clara I, ten Have M, Belik SL, de Graaf R, Stein MB (2007). Perceived barriers to mental health service utilization in the United States, Ontario, and the Netherlands. Psychiatric Services 58, 357–364.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E. ... Dunbar, G. C. (1998). 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. *Journal of Clinical Psychiatry,* 59, 22-33.

Smith, A. (2017). Record shares of americans now own smartphones, have home broadband. Pew Research Center. url: http://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/

Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine. May 22 2006;166(10):1092-1097. PMID: 16717171

Storch, E., Larson, M., Shapira, N., Ward, H., Murphy, T., Geffken, G., . . . Goodman, W. (2006). Clinical predictors of early fluoxetine treatment response in obsessive–compulsive disorder. Depression and Anxiety, 23(7), 429-433.

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015, 2015/03/11). Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR mHealth uHealth, 3*(1), e27. https://doi.org/10.2196/mhealth.3422

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J Exp Psychol, 18(6)*, 643.

Wilhelm, S., & Steketee, G. S. (2006). Cognitive therapy for obsessive compulsive disorder: a guide for professionals. Oakland, CA: New Harbinger Publ.

Wootton, B., Dear, B., Johnston, L., Terides, M., & Titov, N. (2013). Remote treatment of obsessive-compulsive disorder: A randomized controlled trial. Journal of Obsessive-Compulsive and Related Disorders, 2(4), 375-384.

Yovel, I., Gershuny, B., Steketee, G., Buhlmann, U., Fama, J., Mitchell, J., & Wilhelm, S. (2012). Idiosyncratic Severity Profiles of Obsessive Compulsive Dysfunction: A Detailed Self Report Assessment of a Multifaceted Syndrome. Cognitive Therapy and Research, 36(6), 694-713.