Study Protocol and

Statistical Analysis Plan (SAP)

SMARTPHONE COGNITIVE BEHAVIORAL THERAPY FOR BODY DYSMORPHIC DISORDER: A RANDOMIZED, WAITLIST-CONTROL TRIAL

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

Ethical approval:	**Mass General Brigham Institutional Review Boards (protocol # 2017P000293, approved on 04/03/2017)**
ClinicalTrials.gov identifier:	**NCT03673046**
Funding:	**This study was funded by Koa Health (formerly Telefonica Alpha). Investigators from MGH developed the Perspectives 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:	**June 21, 2022 (shortened version)** **Last IRB-approved update: March 11, 2022**

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

Version	**Draft 1.0**

DEVELOPMENT OF A COGNITIVE BEHAVIORAL MOBILE APP FOR BODY DYSMORPHIC DISORDER – PHASE II: TEST OF EFFICACY

I. BACKGROUND AND SIGNIFICANCE

A. Background

BDD is a psychiatric illness that involves a preoccupation with an imagined or greatly exaggerated defect in one's physical appearance (American Psychiatric Association [APA], 2013). This obsessive preoccupation causes clinically significant distress or impairment in functioning (e.g., social, occupational functioning). BDD is also defined by repetitive, compulsive behaviors performed in response to appearance concerns, such as frequent mirror checking (Alliez & Robin, 1969), excessive grooming (Vallat et al., 1971), and skin picking (Phillips & Taub, 1995). BDD is classified as an Obsessive Compulsive Related Disorder in the DSM-5, to reflect its similarities with obsessive compulsive disorder (OCD).

Individuals with BDD often have poor insight, such that the individual fixedly believes that their appearance is truly defective. In fact, approximately one third of patients have delusional beliefs (Phillips, 2006a). Delusions in BDD focus solely on one's appearance, in contrast to severe psychotic illnesses like schizophrenia. Individuals with BDD also experience high rates of comorbid major depression (i.e., in 53-81% of cases) and elevated risk for suicide (Frare et al., 2004; Phillips et al., 2005; Phillips et al., 2007). BDD's early onset and chronicity underscore the need to widely disseminate effective treatments.

Cognitive-behavioral therapy (CBT) is the best-studied and most promising form of psychological treatment for BDD (for meta-analysis see Harrison et al., 2016). CBT for BDD is a time-limited psychotherapy focused on targeting maladaptive thinking and self-defeating behaviors. Specifically, CBT aims to teach patients how to identify and challenge maladaptive thoughts ("cognitive therapy," [CT]), while reducing safety behaviors such as rituals and avoidance ("behavior therapy," [BT]). Most studies have included both cognitive and behavioral components, consisting mainly of exposure and response prevention (ERP) to reduce social avoidance and repetitive behaviors (such as mirror checking) and have shown that CBT is effective for BDD in both group (Rosen, Reiter, and Orosan, 1995; Wilhelm et al., 1999) and individual formats (McKay et al., 1997; Neziroglu et al., 1996; Veale et al., 1996; Wilhelm et al., 2019). Furthermore, one recent study has developed and pilot-tested an Internet-based CBT treatment for BDD (Enander et al., 2016). Participants were randomly assigned to receive either 12 weeks of CBT or supportive psychotherapy for their BDD over the Internet, with a very limited amount of therapist support via the online program. Those in the CBT group showed significantly greater improvement compared to those in the supportive psychotherapy arm, and 56% of those in the CBT group were treatment responders compared to 13% in the supportive psychotherapy arm (Enander et al., 2016). Secondary outcomes such as depression and quality of life also showed significantly greater improvement in the CBT group compared to supportive psychotherapy group, suggesting that technology-based CBT treatments for BDD may be feasible, safe, and effective.

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

While well-validated treatment for BDD exists, CBT for BDD is highly specialized and, thus, very difficult for patients to access. At our specialty program (the MGH BDD Program), there are consistent waitlists of 3-6 months to obtain access to a CBT therapist for BDD. In rural locations, access to this specialized treatment is likely to be even more challenging. Untreated BDD has a chronic, unremitting course, underscoring the importance of access to treatment. Inadequate treatment access due to limited professionals offering this specialized treatment is compounded by economic barriers and shame

preventing many sufferers from seeking in-person care (Fang et al., 2014). Mobile app-based CBT would solve the access gap by addressing each of these barriers.

To this end, development of a high-quality CBT mobile app based on our program's empirically supported intervention is likely to benefit individuals with body image concerns, ranging from the severe end (i.e., BDD), which occurs in 1.7-2.4% of the population (Fang et al., 2014) to milder subclinical body image disturbances, which occur in nearly half the population (Cash & Henry, 1995), by providing low-cost, rapid, widely available access to CBT for BDD. The app may also benefit healthcare providers (e.g., psychologists, psychiatrists, dermatologists) as it could be used as either a stand-alone treatment or in conjunction with treatments with a provider. We recently developed a smartphone-delivered CBT treatment for BDD and examined its preliminary feasibility and acceptability in participants seeking treatment for BDD (see study record NCT03221738). In this trial, we sought to test the efficacy of the smartphone-delivered CBT treatment for BDD.

II. SPECIFIC AIMS

The purpose of this project was to test the efficacy of a mobile-app version of CBT treatment for BDD.

Specific Aim 1: In a waitlist-controlled trial, we will test the efficacy of the active treatment arm compared to a waitlist in reducing primary (i.e., BDD severity) at end-of-treatment (week 12).

> Primary Hypothesis: We hypothesize that participants receiving app-based CBT for BDD will have greater improvement in BDD-YBOCS scores than those in the waitlist condition at treatment endpoint (week 12).

Specific Aim 2: In a waitlist-controlled trial, we will test the efficacy of the active treatment arm compared to a waitlist in reducing secondary outcomes (e.g., delusionality, quality of life, social and occupational functioning, depression) at end-of-treatment (week 12).

> Secondary Hypothesis: We hypothesize that participants receiving app-based CBT for BDD will have greater improvement on secondary clinical outcome measures (i.e., depression (QIDS-SR), delusionality (BABS), functional impairment (SDS), and quality of life (Q-LES-Q)) than those in the waitlist condition at treatment endpoint (week 12).

III. SUBJECT SELECTION

Overview:

Massachusetts General Hospital was responsible for all participant recruitment and enrollment. We collaborated with Koa Health B.V., a in the testing of the app.

For this Phase 2 randomized waitlist control trial, our target sample size was 80 eligible individuals with primary BDD (for details, see section VI-D "Power analysis"). Detailed eligibility criteria for each aim follow. Additionally, a detailed plan for safety and risk management is described below.

Inclusion/Exclusion Criteria:

1. Inclusion criteria
    a. at least 18 years of age
    b. current diagnosis of primary DSM-5 BDD, based on 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 BDD
    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. Concurrent psychological treatment
    h. Does not own a supported mobile Smartphone with a data plan
    i. Lack of technology literacy that would interfere with ability to engage with smartphone treatment

A. Source of Subjects and Recruitment Methods:

Overview:

Participants were informed about the study through advertisements on public transportation, 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, 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 the recruitment website created 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. Individuals were also recruited as part of the BDD clinic's general recruitment protocol. Interested individuals were referred to the study research assistant (RA), who provided more information about the research study and assessed preliminary eligibility over the telephone. We also used a screening questionnaire on REDCap; the link to access this screener was embedded within our recruitment website.

IV. SUBJECT ENROLLMENT

A. Procedures for obtaining informed consent and assessing eligibility

Interested and eligible participants were invited to complete the baseline assessment with a trained independent evaluator (IE) by Healthcare secure Zoom, phone, or Virtual Visit, a Partners IS-approved, secure, HIPAA-compliant video calling program (see more Virtual Visit information below: "Monitoring and Quality Assurance"), as recruitment was nation-wide. At that time, patients were informed about the study's purpose and procedures in more detailed and were advised regarding alternative treatment options. Before eligibility was assessed, trained clinician-level staff obtained electronic informed consent through REDCap, a secure data capture system (see more REDCap information below: "Monitoring and Quality Assurance"). The person obtaining informed consent also verified the participant's identity by requesting to view a form of identification (e.g., government-issued driver's license). Subjects were given as much time as they needed to consider participation. Participants had the ability to download and print or save the electronic informed consent document. The research assistant also emailed participants a signed pdf copy of the consent form for their records.

B. Treatment assignment and randomization (if applicable)

Eligible subjects were randomized to immediate app-based CBT for BDD or to a waitlist control condition (50/50 chance). Randomization was stratified by participants' medication status, in order to evenly distribute any potential medication effects on treatment response across both study arms. The waitlist control group allowed us to determine whether active treatment is superior to gains afforded by the passage of time or other extraneous variables (e.g., history, maturation, testing). Those beginning the CBT treatment immediately were taught to download the app and received a brief tutorial in using the app-based treatment. Following the waitlist period, participants randomized to this condition were crossed over to app-based CBT to allow assessment of within-subject treatment gains for control participants. The duration of CBT for waitlist participants was the same as for participants who were treated first with CBT.

Wait Period: To maximize the validity of our study, we matched the duration of the waitlist condition to that of the CBT condition. We successfully used a similar design in the PI's prior BDD trial assessing in-person individual CBT. Furthermore, Veale et al (1996) has shown that patients do not improve on a 12-week waitlist and that this duration is tolerated by BDD patients. To protect the safety of participants assigned to the waitlist condition, participants who had active suicidal ideation at the screening assessment were not eligible to participate in this trial (see Exclusion criteria). Additionally, individuals with severe comorbid major depression were not eligible to participate this trial (see Exclusion criteria). The waiting period for standard care for BDD at MGH is at least 3-6 months. Thus, being on a 3-month waitlist in the proposed study may enable the subject to start treatment sooner than being on the general clinic waitlist. However, we offered referral resources to study candidates who did not wish to be randomly assigned to a possible waitlist condition.

V. STUDY PROCEDURES

Following screening and randomization, participants completed self-report measures for the baseline assessment through a secure REDCap link emailed to them. In this trial, we aimed to evaluate changes in primary symptoms as well as to 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 depression, functional impairment, delusionality, 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: Phone screen procedures were as described above.

Screening Calls and Randomization Procedures: Procedures were as described above.

Assessments: Participants in both conditions were assessed at regular intervals: at baseline (week 0), mid-treatment (week 6), end-of-treatment (week 12), and a 3-month follow-up visit (for details, see Table 1 below). Masters or doctoral-level study IEs conducted clinician-rated assessments, and participants completed self-report questionnaires on the computer via REDCap, a secure data capture system. Please see the schedule of assessments and measures administered (relevant to the primary and secondary aims of this trial) at each time point in Table 1. Assessments required approximately 1 hour. These assessment lengths are like those used in other studies in our program and have been well-tolerated by participants. To avoid dropout, participants were paid $25 for mid-waitlist, end-of-waitlist, mid-treatment, end-of-treatment, 3-month follow up assessments, and 12-month follow-up assessments. Despite our best efforts to retain participants, we expected dropouts. Except for subjects who withdrew consent to participate, all

who were withdrawn or dropped out of the study were asked to complete all scheduled assessments and were provided remuneration for participation in assessments.

3-month Follow-up Feedback to Technologists: Participants could choose to meet with technologists from Koa Health (sponsor) via a Partners-approved, secure, HIPAA-compliant clinical video platform (Healthcare secure Zoom) at the 3-month follow-up, for the technology experts to obtain input and feedback on the usability and feasibility of the CBT app. This meeting would take approximately 1.5 hours, and participants would be reimbursed $25 for completion of the 3-month follow-up feedback visit with technologists.

Table 1: Measures table

	RANDOMIZED CONTROL TRIAL: ARM 1								COMPASSIONATE CARE (WAITLIST CONDITION CROSSES TO APP)			
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*	Weekly	Week 4	Mid-Treatment Assess (wk 6)	Post-Treatment Assess (wk 12)
Treatment History Checklist	✔											
Treatment History Medication	✔											
Treatment History Psychosocial	✔											
AE Form		✔			✔	✔	✔	✔			✔	✔
Body Parts of Concern Checklist (BPCC)	✔											
BDD Y-BOCS	✔	✔			✔	✔	✔	✔			✔	✔
BABS	✔	✔			✔	✔	✔	✔			✔	✔
MINI	✔											
C-SSRS Lifetime	✔											
C-SSRS Update		✔				†						
CGI–C	✔	✔			✔	✔	✔	✔			✔	✔
Treatment Condition Questionnaire						✔						
CONCOM Form		✔			✔	✔	✔	✔			✔	✔
BDD-PSR	✔	✔			✔	✔	✔	✔			✔	✔
Life Events Question		✔			✔	✔	✔	✔			✔	✔
Technologist Interview Notes							✔					
PHQ-2			✔						✔			
QIDS-SR – Q12			✔						✔			

CGI-P		✔	✔		✔	✔	✔	✔	✔		✔	✔
Demographics	✔											
Lifetime Psychiatric Medications	✔											
Credibility Expectancy Rating~	✔				✔	†					✔	
QIDS-SR	✔	✔			✔	✔	✔	✔			✔	✔
QLESQ	✔				✔	✔	✔	✔			✔	✔
SDS	✔				✔	✔	✔	✔			✔	✔
CSQ~					✔	✔					✔	✔
Treatment Utilization Questionnaire~					✔	✔	✔				✔	✔
App Feedback Questionnaire (Endpoint)~						✔						✔
App Feedback Questionnaire (Baseline)~	✔					†						
App Feedback Questionnaire (Midpoint)~					✔						✔	

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 the AE form, AE log, BDD-YBOCS, BABS, CONCOM form, CMED log, and CTHERAPY log to that participant, and they would complete an additional CGI-P (see "Baseline Re-Assess," above).

† Post-treatment assessment also functions as the baseline assessment for those in waitlist condition, who cross over into compassionate care. † indicates supplemental measures added only for compassionate care baseline.

* Participants in CBT condition go directly into follow-up assessments after week 12. Those in waitlist condition complete follow-up assessments after compassionate care.

~Only administered in the RCT active treatment group (not waitlist) during the first 12 weeks of the trial.

Measures Descriptions

Diagnostic Measures

Mini International Neuropsychiatric Interview (M.I.N.I.): The M.I.N.I. (Sheehan et al., 1998) is a semi-structured diagnostic assessment of DSM-5 psychiatric illnesses. The M.I.N.I. is efficient, reliable, and well-validated. It was the sole/primary diagnostic evaluation used to assess for psychiatric diagnoses and establish inclusion and exclusion criteria.

Assessment of Body Image and Related Symptoms

Yale Brown Obsessive Compulsive Scale Modified for BDD (BDD-YBOCS) : This gold-standard 12-item semi-structured clinician-administered scale rates past-week BDD symptom severity (Phillips et al., 1997). It will be the study's primary outcome measure. The BDD-YBOCS has excellent internal consistency (α=.80), interrater and test-retest reliability (ICC for total score=.99 and .88, respectively), convergent validity (r=.55 with the CGI), and sensitivity to change (Phillips et al., 1997).

BDD-Symptom Scale (BDD-SS): The PI developed the BDD-SS to rate the severity of specific BDD symptoms (Wilhelm et al., 2016). It will be used in the present study to obtain a broader assessment of participants' BDD symptoms and associated thoughts, feelings, and behaviors.

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

Psychiatric Status Rating Scale for Body Dysmorphic Disorder (BDD PSR): The BDD PSR (Phillips et al., 2006b; Phillips et al., 2013) is a 7-item, clinician-completed rating scale of the patient's BDD diagnostic status. It ranges from 1 (Full Remission) to 7 (Meets full diagnostic criteria with extreme/severe BDD).

Brown Assessment of Beliefs Scale (BABS): This 7-item semi-structured clinician-administered interview assesses delusional thinking related to one's appearance concerns (Eisen et al., 1998). It has very strong psychometric properties, including internal consistency, interrater and test-retest reliability, convergent validity, divergent validity, and sensitivity to change (Eisen et al., 1998).

Patient Health Questionnaire-2 (PHQ-2): 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 was administered weekly via the app to monitor changes in depression symptom severity.

Quick Inventory of Depressive Symptomatology- Self Report (QIDS-SR): 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". Scores on the QIDS-SR were used to evaluate eligibility at the screening visit, and to measure self-reported depressive symptoms at subsequent assessment visits. The QIDS-SR i12 was delivered weekly to participants via the app (in the active group) or via REDCap (in the waitlist group) to monitor for risk concerns during the trial. Scores >0 triggered a popup message to the participant about calling 911/going to the ER, and participants were provided information about contacting a suicide hotline. Scores >1 also triggered a text message alert to the coach with the specific item response given by the participant (i.e., 2 vs. 3). Study clinicians followed 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). After an initial alert for a participant, study clinicians used clinical judgement when determining follow-up for additional alerts.

Columbia-Suicide Severity Rating Scale (C-SSRS): The C-SSRS (Posner et al., 2011) is a gold-standard, clinician-administered assessment of suicidal ideation and suicide behaviors. The baseline version, which assesses both lifetime and recent time frames, was used to establish eligibility at the baseline/screening visit. The "Update" version, which assesses suicide risk since the last assessment, was used at subsequent study administrations.

Body Parts of Concern List: This list names a wide range of body parts that are the most common areas of appearance concerns for BDD patients. The measure asks participants to indicate which body parts they are preoccupied with.

Assessment of Functioning and Quality of Life

Sheehan Disability Scale (SDS): The SDS uses a Likert scale from 0 (not at all) to 10 (extremely) to assess impairment in occupational, social, and family domains. It has strong internal consistency and validity (Sheehan et al., 1996).

Quality of Life Enjoyment and Satisfaction Questionnaire – Short Form (Q-LES-Q-SF): 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

Credibility and Expectancy Questionnaire (CEQ): This 6-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): The CSQ is a 25-item self-report questionnaire which assesses the satisfaction with clinical services received. It has excellent internal consistency and good discriminant validity (McMurtry & Hudson, 2000).

Open-Ended Feedback Question: is a 1-item open-ended question to obtain feedback from the participant after completion of each treatment module. The question asks, "Can you describe what you liked and didn't like about this activity?" Participants provide a free-text response.

Other Assessments

Treatment History Medication, Treatment History Psychosocial, Treatment History Checklist forms: These forms together were used to assess any current medications (past 2 months) taken as well as any current or lifetime psychosocial treatment.

Lifetime Psychiatric Medication Form: This self-report form collected lifetime psychiatric medications taken and was administered at the baseline screening visit.

Concomitant medication and therapy form: This log tracked any changes in medication and therapy that the participant made since the prior assessment.

Demographics Form: This self-report form collected basic demographic data and was administered at baseline.

Treatment Utilization Questionnaire: This self-report form measures how much time participants are practicing treatment skills both on and off the app. It was completed as a self-report measure at the mid-treatment assessment and end of treatment visits.

App Feedback Questionnaire (Baseline): 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): These multiple choice/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 Feedback Questionnaire (Endpoint): This self-report form collects participant feedback pertaining to the content and aesthetics of the app (e.g., "How clear was the layout of the app?")

Technologist Interview Notes: This interview guide asks about credibility/expectancy, client satisfaction, and improvements at trial completion. Of note, technologists from Koa Health who conducted over-the-phone feedback interviews were hired and paid by the sponsor, Koa Health. These interviews were not considered part of the human subjects research and was development work for the app. Therefore, they were not included as study staff, but the informed consent process fully informed participants about the interactions with technologists and as such, these procedures are described in full in this protocol. MGH study staff introduced participants to the technology experts via phone call or email introduction after the participant was fully consented. This served to protect participants' identifying information (e.g., we did not share access to participants' email addresses or phone numbers, but rather MGH study staff-initiated conference calls and scheduled appointments between participants and technologists).

Treatment Condition Questionnaire: This brief measure will be used to assess whether the independent evaluator believes each subject was assigned to either immediate CBT or waitlist, as well as their confidence in this belief. This questionnaire was completed at the end-of-treatment assessment (i.e., the week 12 visit). This measure is based on Bang's 2x5 Blinding Index (Bang et al., 2004).

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

App-Based Data

Data from sensors in participants' mobile phones were collected, to optimize the program through personalization and improve the app. (For a detailed description of storage and protection of de-identified mobile data, see Monitoring and Quality Assurance, below.)

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.).

The Treatment

The CBT app included the following components of CBT for BDD: 1) education about a CBT model of BDD; 2) use of self-monitoring to record trigger situations and symptoms; 3) cognitive techniques to identify and challenge distorted thoughts related to one's BDD; 4) exposure to avoided situations; 5) response prevention to decrease repetitive behaviors; 6) mindfulness/ perceptual retraining (to help patients to learn to balance distressing emotional states with rational thinking and to control their attentional processes); 7) increasing valued activities, and 8) relapse prevention (to teach patients to expect and react effectively to setbacks that may occur during times of stress). Details of the app-treatment are described further in the manuscript of the pilot study (Wilhelm et al., 2020).

Treatment Length: We tested a 12-week version of app-treatment. Twelve-week CBT treatments for BDD have been tested previously and shown to be acceptable and effective (e.g., Veale et al., 2014), including in a computer-delivered format (Enander et al., 2016). Participants had access to the Perspectives app and coach support throughout the 12-week treatment. During the follow-up phase of the study, participants continued to have access to the Perspectives app, but without the coach support or messaging system.

Module Frequency and Duration: An advantage of app-based CBT for BDD is that participants were able to self-direct the frequency and duration of modules, and they were able to re-visit sessions as many times as is useful to them.

Treatment Format: Each treatment component was presented through modules on the mobile app, and exercises were logged and practiced through the app on participants' smartphones. Upon initiating the app-based CBT treatment, each participant was assigned a BA-level coach with some training in CBT and some familiarity with BDD; coaches will be supervised weekly by a licensed psychologist (see "Coach Training and Qualifications," below). Participants were able to communicate with their coach through a secure messaging system incorporated into the app throughout the 12-week treatment. The coaches had a separate portal inside the Partners firewall to receive and respond to these in-app communications from patients. The coach communication aimed to provide support and additional motivational enhancement to patients, and to provide feedback about the skills and homework that the patients were learning through the app. Patients were notified that their coach would respond to all in-app communications within 36 hours on weekdays. Moreover, brief phone check-ins could 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). During the 3-month follow-up phase of the study, participants continued to have access to the app, but without the coach support or messaging system. Participants lost access to the app following the 3-month follow-up period and did not have access to the app between the 3-month and 12-month follow-up assessments. Based on a similar, online CBT treatment for BDD (Enander et al., 2016)

and our own feasibility and acceptability trial (Wilhelm et al., 2020), we expected average coach contact per week to be minimal (e.g., less than 15 minutes per week, per patient).

In-App Coach Training and Qualifications: BA-level coaches were study staff members with a Bachelor of Arts (BA) or Bachelor of Science (BS) degree in Psychology or a related field. Based on preliminary findings from our feasibility and acceptability trial, we anticipated that the coaches' involvement would be light-touch and focus mostly on motivation and problem-solving, as needed, whereas the app itself would 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 received training before assisting in treatment as a coach (e.g., completing the MGH Psychiatry Academy CBT training course for BDD) and were required to pass (>90% correct) the MGH Psychiatry Academy BDD knowledge test. To ensure ongoing high-quality treatment, study staff members who were licensed clinicians with expertise in CBT for BDD provided weekly supervision to coaches, with additional supervision as needed if questions arose.

IE Training and Qualifications, and Procedures to Ensure Assessment Integrity and Interrater Reliability

Assessments were conducted by independent evaluators (IE) who had Masters or Doctoral-level degree in clinical psychology or related mental health fields and who were employed at MGH. The IEs were otherwise uninvolved in study procedures. Training and reliability checks were done to ensure that IEs conducted ratings in a uniform way. Raters first received instruction in the SCID-I/P, MINI, SCID-II, BDD-YBOCS, BABS, and CGI from Dr. Wilhelm or another gold-standard expert rater, prior to beginning as an IE. The IEs were then supervised twice monthly by the PI. The IE were required to demonstrate reliability on the BDD-YBOCS and BABS at a criterion of .80 ICC, compared to measures rated by Dr. Wilhelm or another gold-standard rater. All assessments were audiotaped for reliability ratings. Names were not included on digital recordings. To reduce rater drift, a trained reliability rater meeting the same qualifications as the IE reviewed 15% of randomly selected audiotaped interviews at regular intervals. IEs were blind to randomized conditions.

Privacy and Confidentiality

All information gathered was kept strictly confidential. We adhered to the following procedures to protect privacy and confidentiality:

1) Participants were assigned a code number (study ID) that was used in all email communications or shared data files outside of the secure MGH lab server or Partners Dropbox Business. Participants' names or other identifying information did not appear on any questionnaires, study documents, digital recordings, computerized data files, or published reports. Case records were 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) were educated about the importance of strictly protecting participants' rights to confidentiality.

2) All personnel was 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) had access to the identifiable data.

3) Clinical data collected during MGH assessments: Computerized data and digital audio files were stored de-identified, in password protected files saved on the protected lab server. 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. Data servers are stored within PHS IS corporate firewall, in a secure, key access facility with password protected computers. Only trained 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. Hard copy data (paper forms) were stored securely in locked file cabinets, within locked offices in the MGH OCD and Related Disorders Program.

4) Technologist data: Computerized data and digital audio files collected from feedback interviews with technologists were stored de-identified, in password protected files in the Partners Healthcare Dropbox Business folder.

5) Mobile data: All mobile data was transmitted from the participant's mobile device to the clinical study server via an encrypted Internet connection. The clinical study server was located in the PHS/MGH Secure Data Center. The collected data was linked to the study identifier only. Prior to deployment of the digital application, the application was scanned with Veracode.

6) Telehealth assessments were conducted using the Partners-approved platform for clinical video calls, which has been vetted by Partners to be secure and HIPAA-compliant. Participants were 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 is shared with the sponsor (Koa Health), under the terms of the Data Use Agreement (i.e., Statement of Work).

B. Drugs to be used

Not applicable

C. Devices to be used

Participants downloaded the CBT for BDD app onto their personal smartphone devices. The CBT for BDD 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 RA 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."

The Center for OCD and Related Disorders biostatistician was responsible for data management and analysis. All project staff received training in data management and data confidentiality procedures. Data checks were done regularly to assure that all forms were entered and available for analysis. Data and

analysis files were be backed up on the lab server and, if paper records were used, were stored in separate locked cabinets.

To characterize our sample, data was displayed graphically, and summary statistics (e.g., means and frequencies) were 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 were 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 were screened for inconsistent or abnormal values. Continuous measures were assessed for skewness and outliers (based on model residuals); no data transformations were necessary to meet modeling assumptions of normality and homogeneity of variance. A two-tailed p-value <.05 was considered evidence of statistical significance for the primary and secondary outcomes. Our primary and secondary analyses used 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 expected dropouts (subjects who are lost to follow-up or withdraw early from the study). We attempted to perform all scheduled assessments for subjects who were withdrawn from the protocol and provided financial incentives for participation in assessments. Our intent-to-treat sample included all randomized patients who completed at least one assessment, making our primary and secondary analyses intent-to-treat analyses. We also performed "per protocol" analyses. Our per protocol population included all randomized patients who completed the baseline and post-treatment assessments, and who did not initiate prohibited treatment during the study. We repeated our GLMM analyses using the per protocol sample. Dropout, study withdrawal, and loss to follow-up was tabulated by reason and treatment arm. Analyses were conducted to ascertain to what extent dropouts were nonrandom, and, if so, what factors were associated with dropout. To evaluate potential attrition bias, study dropouts were compared to study completers.

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

B. Study endpoints

Primary endpoint: Difference in BDD severity (BDD-YBOCS) at the end of treatment/waitlist period.
Hypothesis 1: Participants receiving app-CBT will have greater improvement in BDD-YBOCS scores than those in the waitlist condition at treatment endpoint (week 12).

Secondary endpoints: Difference in secondary clinical outcomes at the end of treatment/waitlist period
Hypothesis 2: Participants who receive app-CBT will have greater improvement on secondary clinical outcome measures (i.e., depression (QIDS-SR), delusionality (BABS), functional impairment (SDS), and quality of life (Q-LES-Q)) than those in the waitlist condition at treatment endpoint (week 12).

C. Statistical methods

Primary endpoint: Preliminary efficacy for reducing BDD symptom severity
Analysis:

The primary outcome model was a hierarchical mixed model (i.e., GLMM) that included time (categorical; baseline, mid-point, and end-of-treatment), treatment (app-CBT vs. waitlist), and their interaction as fixed effects, and modeled 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 main hypothesis test was based on a specific contrast of treatment difference at week 12. Because these analyses included all subjects who were randomized and completed at least one assessment, it was an intent-to-treat analysis. These analyses included only the 3 assessments from baseline to post-treatment and did not include the follow-up (FU) assessments, because treatment response could be affected by subjects seeking other treatments during the FU period. Between-group effect sizes were calculated using Cohen's *d*. This analysis was then repeated using the per-protocol sample.

In addition to the analyses specified above, we described the proportion of subjects in each treatment arm who achieved response, where response was defined as a 30% or greater reduction in BDD-YBOCS score from baseline to end of treatment (week 12). This empirically derived cut-off of 30% reflects clinically significant improvement and is widely used in BDD treatment studies (Phillips et al., 1997). While no formal comparison was made between treatments, the reporting of response rates is clinically useful and will facilitate comparison to other studies in the field. We reported response as observed (excluding subjects with missing data at week 12).

Secondary endpoint: Preliminary efficacy for secondary outcomes

Analysis:

We used the same hierarchical mixed model approach as described for the primary endpoint to examine significant differences in secondary outcome measures (i.e., depression severity, delusionality, functional impairment, and quality of life) at posttreatment for treated participants compared to waitlist control participants. We did 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.

D. Power analysis

Pre-COVID-19, the sample size required to test a significant symptom severity difference between the app-CBT and the waitlist control conditions at the end of treatment was estimated to be n=64 (n=32 per group). The sample size estimates were 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 (app-CBT vs. waitlist), and an anticipated effect size of 0.9. We further assumed a drop-out rate of 15% by the end of treatment. This anticipated effect size of the app-CBT is in the range of effect sizes calculated using the published estimate of a waitlist control group BDD-YBOCS mean after 12 weeks (source: Wilhelm et al., 2014), compared to either the combined site mean at week 12 in an in-person CBT treatment study for people with BDD (*d*=0.93; Wilhelm et al., 2019) or the mean at week 12 in an internet-based CBT treatment study for people with BDD (*d*=1.02; Enander et al., 2016). More generally, large effect sizes for CBT vs. waitlist are to be expected, based on meta-analysis by Ost and colleagues (2015) for the closely related disorder OCD, which found that the weighted mean effect size of CBT vs. waitlist control was Hedges *g*=1.31, 95%CI: [1.08, 1.55]. The anticipated drop-rate was based on the weighted mean of drop-out rates (14.4%) reported in three studies that examined the efficacy of CBT treatment in participants with BDD: one compared to waitlist (n=36, 19.4% drop-out at week 12; Wilhelm et al., 2014), one compared to SPT,

in person (n=120, 23.3% drop-out at week 24; Wilhelm et al., 2019), and one comparing internet-based CBT vs. SPT (n=94, 1.1% drop-out at week 12; Enander et al., 2016).

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 - were likely to increase in the US population. Furthermore, the variability of these negative mental health states and stressors would have likely increased as multiple waves of the pandemic and state-wide measures of social interaction restrictions swept the nation in the foreseeable future. In response, we increased the planned sample size to a total n of 80 (40 per group) of participants who start treatment (i.e., those who are eligible and willing and able to participate); this increase was intended to be reflective of a decrease in the anticipated effect size from 0.90 to 0.81 (i.e., a 10% decrease of the anticipated effect size). The plan for and justification of the sample size increase was reviewed by the Institutional Review Board at MGB prior to implementation.

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 or supporting clinician/coach dashboard malfunctions or does not work for a period of time, the patient may be unable to use the app to receive treatment or communicate with their coach. Similarly, if the clinician/coach dashboard were inaccessible, the coach would not be able to communicate via the app 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 were 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, were 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) were 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 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 or suicidal ideation related to the natural waxing and waning of BDD or other psychiatric (e.g., depression) symptoms. BDD can be associated with other psychiatric symptoms (e.g., depression, anxiety), as well as suicidal thoughts and behaviors, which may also change over time. Breach of confidentiality represents a potential risk. Finally, the treatment procedures, particularly the exposure

exercises, will potentially provoke some anxiety. As discussed below, we took precautions to ensure that these potential risks were minimized (see Adequacy of Protection Against Risks below).

Minimizing of Risks and Safety Reporting.

The following procedures were implemented to protect participants against risks:

1. Participants with active suicidal ideation at the screening assessment were excluded from participating (see Inclusion/Exclusion criteria). If a subject scored ≥21 on the QIDS-SR and/or >1 on the QIDS-SR item #12, the independent evaluator followed up with a phone evaluation within 24 hours and referred the participant to a higher level of care if clinically indicated. Suicidal ideation at screening was also measured using the C-SSRS (see Inclusion/Exclusion criteria), and if a potential participant scored ≥2, a study clinician, PI, or independent evaluator conducted a risk evaluation with the participant and referred the participant to a higher level of care if clinically indicated.

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

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

4. A general resources page was available on the app at all times to participants, which included a suicide hotline number.

   Participants' clinical improvement or deterioration was assessed weekly via a participant-rated CGI-BDD collected via the app (or REDCap, in waitlist condition). Participants could be withdrawn from the study if their clinical condition deteriorated substantially. Deterioration was defined by a combination of (1) a rating of 6 (much worse) or 7 (very much worse) on the weekly, participant-rated CGI-BDD across 2 subsequent weeks and (2) PI judgment that remaining in the study was not in the participant's best interest. Of note, a single rating of 6 or 7 on the weekly, participant-rated BDD-CGI also triggered a notification to the clinician/coach via text message and email. In the case that a BA-level coach was notified that a participant's CGI indicates deterioration, they notified a doctoral-level clinician as soon as possible (and within 24 hours). A study clinician then followed up with a phone evaluation within 24 hours of the alert and referred the participant to a higher level of care if clinically indicated. After an initial alert for a participant, study clinicians could use clinical judgement when determining follow-up for additional alerts.

5. Ratings on the QIDS-SR item 12 (suicide item) were carefully monitored weekly via the app (or REDCap, in waitlist condition); a score >0 at any assessment triggered a pop up message to be presented to the patient within the mobile app reminding participants that if they were experiencing suicidal thoughts, they should seek professional help or go to the emergency room right away. Links to 911, as well as a national suicide hotline, were provided within this popup notification. A score >1 also triggered notification to the clinician/coach via text message or email. In the case that a BA-level coach was notified that a participant reported a score>1, they notified a doctoral-level study clinician as soon as possible (and within 24 hours). A study clinician, PI, or independent evaluator then followed up with a phone evaluation within 24 hours of the alert and referred the participant to a higher level of care if clinically indicated. After an

initial alert for a participant, study clinicians could use clinical judgement when determining follow-up for additional alerts.

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

7. Treatment through app-based CBT was supplemented with electronic communication with a study coach, who could answer questions and guide participants through the treatment as needed. See *In-App Coach Training and Qualifications*, above, for details.

8. The independent evaluator(s) were highly experienced, highly trained, and closely supervised.

9. Dr. Wilhelm (the PI) was available, if necessary, to discuss the study, alternative treatments, or any concerns about the study with participants if requested by the participant, coaches, or raters.

10. The PI and her co-investigator and the participants' study coach were available to answer study questions via the app or phone. This was clearly communicated orally and in writing to study participants.

11. All participants were provided with referral resources.

12. The study coaches and raters (IEs) made every attempt to help participants feel comfortable when discussing sensitive material. Participants could skip questions on assessments that they were uncomfortable answering.

13. If exposure exercises suggested through the treatment app were too anxiety provoking, participants were able to do alternative exercises that caused less anxiety.

14. The CBT treatment initially emphasized cognitive restructuring, which we anticipated would be less anxiety provoking than exposure treatment alone and would make exposure more tolerable.

15. Technologists from Koa Health who conducted the technology feedback interviews were trained professional staff with experience conducting patient interviews. They received additional training (including a multi-day workshop) from our team (MGH OCD and Related Disorders Program) on CBT and issues of confidentiality. They completed CITI training.

16. Three clinical psychologists or researchers familiar with BDD were selected to serve as a Data Safety Monitoring Board, to review the study once a year.

17. Participants designated a relative or friend who could be contacted should the subject be unavailable and the investigator had concerns about the subject's well-being.

We anticipated that the above procedures would 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 BDD expert, from careful clinical monitoring, and by experiencing some relief from their BDD symptoms through the CBT app.

B. Potential benefits to society
If app-based CBT for BDD were efficacious, it would offer increased, cost-effective access to CBT for BDD, a treatment which is otherwise difficult to access.

IX. MONITORING AND QUALITY ASSURANCE

A. **Sponsor-initiated monitoring**
Decus Biomedical, Inc. were hired by Koa Health to serve as external study monitors. They visited MGH to provide feedback regarding ongoing study procedures and data quality.

B. Independent monitoring of source data
The PI had overall responsibility for study data and participant safety. All aspects of the study were conducted in accordance with the hospital's policy on confidentiality.

Paper research records were kept de-identified, in a locked file in a locked office at MGH. Self-report measures and some clinician administered measures were 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. The iterative development and testing process results in a well-planned data collection strategy for individual studies. Web-based self-report measures rely on a study-specific data dictionary built by members of the research team. Once built, participants and study staff can enter data directly into REDCap surveys via any computer or tablet with standard web access and browsers. Participants entered survey responses directly into an electronic assessment form on subject-facing REDCap, and the responses were then be transmitted and stored in the secured and confidential database. Each participant only had access to his or her own survey, but not any other survey data. All data was collected and stored in compliance with the Standards for Privacy of Individually Identifiable Health Information ("Privacy Rule") of the Health Insurance Portability Act of 1996 (HIPAA). Any data that was transmitted electronically was encrypted and password protected. 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. All users had defined roles and privileges pre-determined by the system administrator. Thus, the PI could determine the level of access for each study staff such that only a limited number or people had access to sensitive study data.

We anticipated that the above procedures would ensure the confidentiality and integrity of study data.

C. Safety monitoring
Three clinical psychologists or researchers knowledgeable about BDD were selected to serve as a Data Safety Monitoring Board, to review the study once a year when the study was actively enrolling. The PI had 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.**"

D. Outcomes monitoring

Adverse events and data completeness were monitored regularly throughout the trial as described above.

E. Adverse event reporting guidelines

Adverse event reporting:
Adverse events were reported per PHRC guidelines.

X. REFERENCES

Alliez, J., & Robion, M. (1969). Aspects psycho-pathologiques de la defiguration leur relation avec dysmorphophobie [Psychopathologic aspects of disfiguration; their relation to dysmorphophobia]. *Ann Med Psychol (Paris),* 2, 479-94.

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

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

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

Cash, T.F., & Henry, P.E. (1995). Women's body images: The results of a national survey in the U.S.A. *Sex Roles*, 33, 19-28.

Cook, R. J., & Farewell, V. T. (1996). Multiplicity considerations in the design and analysis of clinical trials. *J R Stat Soc Ser A Stat Soc*, 159(1), 93-110.

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

Eisen, J. L., Phillips, K. A., Baer, L., Beer, D. A., Atala, K. D., & Rasmussen, S. A. (1998). The Brown Assessment of Beliefs Scale: reliability and validity. *Am J Psychiatry*, 155, 102-8.

Enander, J., Andersson, E., Mataix-Cols, D., Lichtenstein, L., Alstrom, K., Andersson, G., Ljotsson, B., & Ruck, C. (2016). Therapist guided Internet based cognitive behavioural therapy for body dysmorphic disorder: Single blind randomised controlled trial. *BMJ,* 352, i241.

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

Fang, A., Matheny, N. L., & Wilhelm, S. (2014). Body Dysmorphic Disorder. *Psychiatr. Clin. North Am,* 37, 287-300.

Frare, F., Perugi, G., Ruffolo, G., & Toni, C. (2004). Obsessive–compulsive disorder and body dysmorphic disorder: a comparison of clinical features. *Eur. Psychiatry*, 19(5), 292-298.

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.

Hamer, R.M., &Simpson, P.M. (2009). Last observation carried forward versus mixed models in the analysis of longitudinal psychiatric clinical trials. *Am J Psychiatry*, 166, 639-641.

Harrison, A., Fernandez de la Cruz, L., Enander, J., Radua, J., & Mataix-Cols, D. (2016). Cognitive-behavioral therapy for body dysmorphic disorder: A systematic review and meta-analysis of randomized controlled trials. *Clin Psychol Rev*, 48, 43-51.

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

McKay, D., Todaro, J., Neziroglu, F., Campisi, T., Moritz, E. K., & Yaryura-Tobias, J. A. (1997). Body dysmorphic disorder: A preliminary evaluation of treatment and maintenance using exposure with response prevention. *Behav Res Ther,* 35, 67-70.

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

Neziroglu, F. (1996). Effect of cognitive behavior therapy on persons with body dysmorphic disorder and comorbid axis II disorders. *Behav Ther,* 27*,* 67-77.

Öst, L. G., Havnen, A., Hansen, B., & Kvale, G. (2015, Aug). Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. *Clin Psychol Rev*, 40, 156-169.

Phillips, K. A., Coles, M. E., Menard, W., Yen, S., Fay, C., & Weisberg, R. B. (2005). Suicidal ideation and suicide attempts in body dysmorphic disorder. *J Clin Psychiatry*, 66(6), 717-725.

Phillips, K. A., Didie, E. R., & Menard, W. (2007). Clinical features and correlates of major depressive disorder in individuals with body dysmorphic disorder. *J Affect Disord*, 97(1), 129-135.

Phillips, K. A., Menard, W., Pagano, M. E., Fay, C., & Stout, R. L. (2006a). Delusional versus nondelusional body dysmorphic disorder: clinical features and course of illness. *J Psychiatr Res*, 40(2), 95-104.

Phillips, K. A., Menard, W., & Fay, C. (2006b). Gender similarities and differences in 200 individuals with body dysmorphic disorder. *Compr Psychiatry*, 47(2), 77-87.

Phillips, K. A., Hollander, E., Rasmussen, S. A., Aronowitz, B. R., DeCaria, C., & Goodman WK. (1997). A severity rating scale for body dysmorphic disorder: Development, reliability, and validity of a modified version of the Yale-Brown Obsessive Compulsive Scale. *Psychopharmacol Bull,* 33(1), 17-22.

Phillips, K.A., Menard, W., Quinn, E., et al. (2013). A 4-year prospective observational follow-up study of course and predictors of course in body dysmorphic disorder. *Psychol Med*, 43, 1109-1117.

Phillips, K. A., & Taub, S. L. (1995). Skin picking as a symptom of body dysmorphic disorder. *Psychopharmacol Bull,* 31, 279-88.

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., Currier, G. W., Melvin, G. A., Greenhill, L., Shen, S., & Mann, J. J. (2011, Dec). The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry*, 168(12), 1266-1277.

Rosen, J. C., Reiter, J., & Orosan, P. (1995, Apr). Cognitive-behavioral body image therapy for body dysmorphic disorder. *J Consult Clin Psychol*, 63(2), 263-269.

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. (2003). 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. *Biol. Psychiatry,* 54, 573–583.

Sheehan, D. V., Harnett-Sheehan, K., & Raj, B. A. (1996). The measurement of disability. *Int Clin Psychopharmacol,* 11, 89–95.

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. *J. Clin. Psychiatry,* 59, 22-33.

Vallat, J. N., Leger, J. M., Destruhaut, J., & Garoux, R. Dysmorphophobias, syndrome or symptom? *Ann Med Psychol (Paris),* 2, 45-65.

Veale, D., Anson, M., Miles, S., Pieta, M., Costa, A., & Ellison, N. (2014). Efficacy of cognitive behaviour therapy versus anxiety management for body dysmorphic disorder: A randomised controlled trial. *Psychother Psychosom,* 83, 341-353.

Veale, D., Gournay, K., Dryden, W., Boocock, A., Shah, F., Willson, R., & Walburn, J. (1996). Body dysmorphic disorder: A cognitive behavioural model and pilot randomised controlled trial. *Behav Res Ther,* 34*,* 717-29.

Wilhelm, S., Greenberg, J. L., Rosenfield, E., Kasarskis, I., & Blashill, A. J. (2016). The Body Dysmorphic Disorder Symptom Scale: Development and preliminary validation of a self-report scale of symptom specific dysfunction. *Body Image*, 17, 82-87.

Wilhelm, S., Phillips, K. A., Didie, E., Buhlmann, U., Greenberg, J. L., Fama, J. M., Keshaviah, A., & Steketee, G. (2014). Modular cognitive-behavioral therapy for body dysmorphic disorder: A randomized controlled trial. *Behav Ther*, 45(3), 314-327.

Wilhelm S, Phillips KA, Greenberg JL, et al. (2019) Efficacy and Posttreatment Effects of Therapist-Delivered Cognitive Behavioral Therapy vs Supportive Psychotherapy for Adults With Body Dysmorphic Disorder: A Randomized Clinical Trial. *JAMA Psychiatry*. 76(4), 363.

Wilhelm, S., Otto, M. W., Lohr, B., & Deckersbach, T. (1999). Cognitive behavior group therapy for body dysmorphic disorder: A case series. *Behav Res Ther,* 37*,* 71-5.

Wilhelm, S., Weingarden, H., Greenberg, J. L., McCoy, T. H., Ladis, I., Summers, B. J., Matic, A., & Harrison, O. (2020, Jan). Development and pilot testing of a cognitive-behavioral therapy digital service for body dysmorphic disorder. *Behav Ther*, 51(1), 15-26.