

Study Protocol

Version 3.1

Key
version 0.x = prior to ethics submission
version 1.x = at and during initial ethics submission & approval process
version 2.x = subsequent to ethics approval (e.g. amendments)
version 3.x = MHRA-related amendments

TITLE
STOP - Successful Treatment of Paranoia: Replacing harmful paranoid thoughts with better alternatives

NAMES PROTOCOL CONTRIBUTORS
Jenny Yiend[1], Che-Wei Hsu[1], Thomas Kabir[2], Jeroen Keppens[3], Philip McGuire[1], Elias Mouchlianitis[1], Emmanuelle Peters[4], Tanya Ricci[1], Sukhwinder Shergill[1], Daniel Stahl[5], George Vamvakas[6], Pamela Jacobsen[7]

Trial registration: ISRCTN17754650

KEYWORDS
Cognitive Bias Modification, paranoia, psychosis, computerized therapy, RCT, interpretation bias, persecutory delusions

ADMINISTRATIVE INFORMATION

Note: the numbers in curly brackets in this protocol refer to SPIRIT checklist item numbers. The order of the items has been modified to group similar items (see http://www.equator-network.org/reporting-guidelines/spirit-2013-statement-defining-standard-protocol-items-for-clinical-trials/).

Title {1}	STOP - Successful Treatment of Paranoia: Replacing harmful paranoid thoughts with better alternatives
Trial registration {2a and 2b}	ISRCTN17754650
Protocol version {3}	3.1
Funding {4}	This research was funded by the Medical Research Council (MRC) Biomedical Catalyst: Developmental Pathway Funding Scheme (DPFS), MRC Reference: MR/V027484/1

Author details {5a}	Jenny Yiend[1], Che-Wei Hsu[1], Thomas Kabir[2], Jeroen Keppens[3], Philip McGuire[1], Elias Mouchlianitis[1], Emmanuelle Peters[4], Tanya Ricci[1], Sukhwinder Shergill[1], Daniel Stahl[5], George Vamvakas[6], Pamela Jacobsen[7] [1] Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [2] The McPin Foundation, London, UK. [3] Department of Informatics, King's College London, London, UK. [4] Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [5] Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [6] Clinical Trials Unit, King's College London, London, UK. [7] Department of Psychology, University of Bath, Bath, UK.
Name and contact information for the trial sponsor {5b}	**Lead Sponsor** Professor Reza Razavi Vice President & Vice Principal (Research) King's College London Room 5.31, James Clerk Maxwell Building 57 Waterloo Road London SE1 8WA Tel: +44 (0)207 8483224 Email: reza.razavi@kcl.ac.uk **Co-Sponsor** Mr Dunstan Nicol-Wilson South London and Maudsley NHS Foundation Trust R&D Department Room W1.08 Institute of Psychiatry, Psychology & Neuroscience (IoPPN) De Crespigny Park London SE5 8AF Tel: +44 (0)207 8480339 Email: slam-ioppn.research@kcl.ac.uk
Role of sponsor {5c}	Neither the Funder nor the Sponsor will have a role in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication.

INTRODUCTION

Background and rationale {6a}

Psychosis is one of the most disabling mental health conditions, with a lifetime rate of 3.5 [1]. It is associated with significant distress, increased unemployment, suicidal ideation, and impaired social functioning and physical ill-health [2]. Persecutory delusions—characterised by paranoid thinking—are the most frequent and clinically significant symptoms of psychosis. Researchers have shown that paranoid thinking in the general population is

continuously distributed, indicating a hierarchical structure to paranoia, ranging from social evaluative concerns (e.g. fears of rejection) to severe threat (e.g. being subject to significant harm) [3-7]. Persecutory delusions fall at the extreme point on the continuum of paranoid belief. As such, they are associated with more distress than other types of delusion [8], are most likely to be acted upon [9], and represent a strong predictor of hospitalization [10]. Over one-third of all UK psychiatric patients suffer from persecutory delusions, often appearing in a range of psychopathologies, including depression [11], bipolar disorder [12], posttraumatic stress disorder [13], anxiety [14], and with the highest prevalence and greatest intensity in schizophrenia [15].

National Institute for Health and Care Excellence (NICE) clinical guidelines recommend using Cognitive Behavioural Therapy (CBT) for treating psychosis. However, CBT is received by only 1 in 10 of those who could benefit and has shown only moderate effect sizes for delusions [16, 17]. Furthermore, a significant proportion of patients suffering from persecutory delusions continue to experience distressing symptoms following psychological treatment [18, 19]. Thus, there is a need for novel, highly accessible, and low-cost interventions for delusions, either as standalone treatments, or as adjuncts to boost existing therapies. New directions in treatments for delusions emphasize briefer, targeted interventions, with a focus on putative causal factors such as cognitive biases [20, 21].

'Cognitive bias' refers to the selective processing of information that matches the core content of the pathology of a disorder (see [7], on a discussion on comparing cognitive deficits and cognitive biases). For example, individuals with threat-related persecutory delusions may 1) repeatedly attend to a certain type of maladaptive stimulus, in preference to more benign alternatives (i.e., attention bias) or 2) consistently interpret emotionally ambiguous information in one (maladaptive) direction (i.e., interpretation bias). Emotional ambiguity, such as a stranger's stare, or a whispering colleague, could evoke either a benign or harmful interpretation (e.g., is the stranger admiring or hostile?; the colleague conspiratorial or playful?). Biased interpretation occurs when there is a consistent pattern of assuming one type of interpretation over another. In psychosis, repeated paranoid interpretations, such as suggested above, could produce an increased perception of danger of harm to the self. This, in turn, may reinforce paranoid beliefs, maintain symptoms, and increase distress about being at risk of observation or harm from others, further promoting biased processing in a self-perpetuating cycle. In theoretical models of psychosis, such biases are proposed as contributors to the formation and maintenance of positive symptoms [8].

Cognitive Bias Modification (CBM) techniques are a theory-driven treatment development that use a computerised task to manipulate biased interpretations and promote more adaptive processing. CBM-pa is a computer desktop version of this class of intervention specifically designed to target paranoid interpretations. Feasibility testing of a six-session CBM-pa has shown positive results [22]. Here, we aim to build on CBM-pa by testing a novel, entirely self-administered digital therapeutic for paranoia called STOP ('Successful Treatment of Paranoia'). STOP is the successor to CBM-pa and takes the form of a 12 session mobile app, adding six newly created sessions. Based on the feedback gleaned from the feasibility study of CBM-pa [23], STOP includes illustrations with each item used in the study (i.e., graphics depicting the ambiguous scenarios and the non-paranoid interpretation that runs counter to the paranoid reader's initial assumption). Furthermore, STOP mitigates adherence issues commonly faced by online interventions by monitoring and rewarding participants with gamification techniques and scheduling protocols. Specifically, clinical researchers will collaboratively work with participants to schedule interventions on the STOP mobile App; they will also have direct contact with participants during main assessment sessions or when participants miss a session.

Identification and Description of the Investigational Device

a) **Summary Description of the Device**. STOP is a mobile app for either android or iOS platforms that delivers standalone Cognitive Bias Modification for Interpretation therapy (CBM-I), using content specifically relevant to paranoia. The app is a 12-week prescription digital therapeutic for individuals with clinically mild to moderate symptoms of paranoia. The app delivers one therapeutic session per week, each

session comprising 40 training items and taking approximately 40 minutes to complete. Training items are short written scenarios, illustrated with sound and graphics, that participants read. The scenarios are ambiguous with possible paranoid interpretations. Users complete missing words and answer questions in a way that encourages non-paranoid, alternative, adaptive beliefs about themselves and others. Session content is interspersed with factual trivia and badges are used to provide motivation and enhance engagement.

b) **Manufacturer:** Avegen Ltd., HQ King's Cross, 344 - 354 Grays Inn Rd, London WC1X 8BP

c) **Device name and version identifier**. STOP (Successful Treatment Of Paranoia). Version for use in this clinical trial is V1.0.

d) **Traceability** is achieved via assignment of software version numbers in the form version 1.x (V1.x), where x stands for any possible future updates to the app (for example performance updates). For Post-Market surveillance, a UDI number will be applied to the STOP application. However, for the early stage (Clinical Investigations), no UDI's shall be employed, just software version numbers. These are to be displayed in the device label within the app that can be accessed from the app menu.

e) **Intended Purpose**: STOP is a standalone software medical device intended to provide Cognitive Bias Modification for Interpretation therapy (CBM-I), using content specifically relevant to paranoia as an adjunct to standard care for patients of 18 years and older, who exhibit clinically mild to moderate symptoms of paranoia and are currently receiving outpatient treatment under the care of a responsible clinician (psychiatrist, clinical psychologist, or other designated keyworker from a multi-disciplinary team, or general practitioner).

f) **Indications for Use:** STOP is indicated for use with patients 18 years of age and older who are currently receiving treatment under the care of a responsible clinician (psychiatrist, clinical psychologist, or other designated keyworker from a multi-disciplinary team, or general practitioner). STOP is indicated as a 12-week prescription-only treatment for patients with mild to moderate clinical symptoms of paranoia, who self-report paranoid thoughts, whose primary language is English and who are not at high risk of suicide. STOP should be used in a quiet, private place where there are no distractions.

Contraindications

- Patients with a history of substance misuse (alcohol or drugs)
- Patients at high risk of suicide (for example suicidal behaviour in the past 1-3 months such as actual attempt, interrupted attempt, aborted attempt, or preparatory acts or behavior)
- Not to be prescribed to users unsuitable or unable to use a digital therapy

Warnings

- If screen flicker is detected, please cease use of STOP and contact Avegen support helpline.
- If negative feelings or symptoms worsen as a result of using this app for more than a day, please contact your support team and cease use of the STOP app until advised further.

- Please consult with your clinician before adopting new behaviours in public settings or if you feel negative or positive changes in your emotions.
- Please keep your login credentials (PIN) private and do not leave your device unattended or share it with others.

Do not cease any existing clinical treatments without first consulting with your clinician. **Cautions**

- STOP does not represent a substitution for a patient's medication or other prescribed therapy. Patients should continue to take their medications and therapies as directed by their healthcare practitioner.
- STOP cannot be used to communicate with a clinician or any other healthcare professional involved in the user's care.
- Participants who are considered to be highly suggestible may need additional clinical oversight as a check against overly trusting behaviour.
- STOP should be used in a quiet, private place where there are no distractions.
- STOP requires patients to have access to an Android or iOS smartphone or tablet
- STOP works best on a minimum screen size of 5.8 inches (14.732 cm). • The clinical benefits of STOP have not been tested beyond 24 weeks • The clinical benefits of STOP on other psychological symptoms (e.g. anxiety, depression) have not been tested
- The ability of STOP to prevent potential relapse after treatment discontinuation has not been studied.

g) **Description of the device** – see (a).

h) **Necessary training and experience**. Useability testing showed that users were able to engage appropriately with the app without specific additional instructions. Nevertheless, users are provided with a full IFU and a Quick User Guide which they are advised to read prior to first time use of the STOP app. In the clinical trial this will be completed with researchers' assistance. In addition, there is an in-app 'Help' section.

i) No specific medical or surgical procedures are involved in use of the device.

j) **References to IB and IFU**:
303876-STOP_Investigator's Brochure
3037876-STOP_Instructions_for_Use

Objectives {7}
The primary objectives are:

1. To test a newly created 12-session visually engaging STOP mobile App by extending its predecessor, CBM-pa, and implementing user feedback gleaned from the CBM-pa feasibility study.

2. To assess the safety and performance of the STOP mobile App

3. To evaluate STOP's effectiveness at different doses, using a primary outcome measured at 24 weeks post-randomisation, by conducting a three-arm double-blind RCT, comparing 6 and 12 sessions of STOP with a 12-session text reading control.

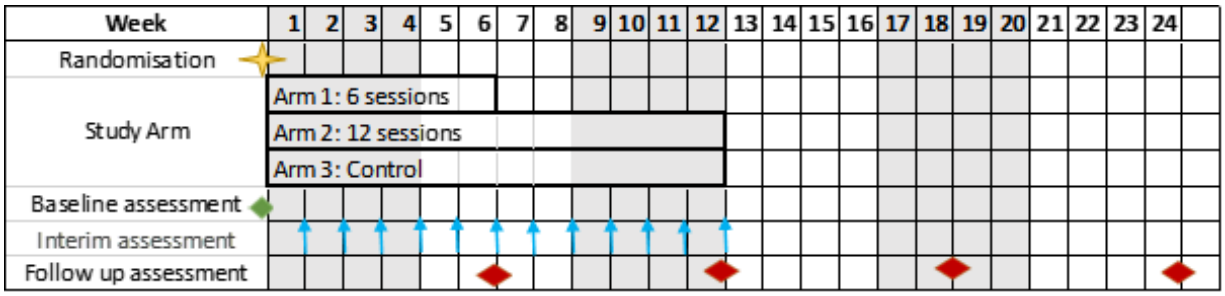
Secondary objectives are:

4. To examine the dose-response relationship by performing a retrospective time-series dose-response modelling of change over time on key outcome measures (a small subset of 4 measures will be completed by patients within the app after each session for this purpose)

Trial Design {8}

We will conduct a four year, two-site double-blind randomised controlled trial (RCT) with patients who experience clinical levels of paranoia. The trial comprises a three arm RCT, comparing two doses of therapy (6-sessions and 12-sessions) with a 12-session text reading control. We will randomly allocate participants into one of the three trial arms: the 6-session intervention group; the 12-session intervention group; or the 12-session text reading control.

Figure 1. Study design



Assessments are given at baseline, 6, 12, 18, and 24 weeks (calculated as post-randomisation), with the primary outcome (severity of paranoid symptoms as measured by the Paranoia Scale (PS) [24] set at 24 weeks. Follow up is anchored to randomisation. The use of a 6-week assessment permits an end of treatment assessment in Arm 1 (6 sessions) and a mid-treatment assessment in Arm 2 (12 sessions). The use of an 18-week assessment permits a 6-week and 12-week ***post-treatment*** comparison between all three arms. The sessional assessments (shown as blue arrows in Figure 1) permit additional, secondary, dose-response analyses alongside the traditional three arm comparisons.

Treatment as usual. All three conditions will be conducted in addition to Treatment as Usual (TAU), which will be in the form of individualised combinations of medication and care coordination and may include certain psychological therapies. We will record details of TAU at each main outcome assessment during the study period using a standardised template to record participants' responses to a series of set questions about treatment received, including pharmacotherapy and any treatment changes. In addition, we will record details of any changes to participants' usual treatment during routine phone calls from researchers using the same standardised template.

Qualitative parallel study. A clinical trainee will conduct qualitative interviews with patients, and focus groups with clinicians, to evaluate whether STOP meets the Tier C

(Interventions) evidence criteria of the NICE Evidence Standards Framework for Digital Health Technologies. Particular focus is around acceptability with users, relevance to care pathways and credibility with health and social care professionals. The trainee's full protocol for this study is given in Appendix 3. Eligibility criteria will be identical to the main trial, except that inclusion criterion 2 (≥-2 on the SRT interpretation bias screening score) will be replaced by 'self-reported paranoid interpretations'. This more accurately reflects the likely access pathway to STOP if delivered in routine clinical care. The samples taking part in the main clinical trial and the qualitative study will be mutually exclusive, meaning that qualitative feedback will be based only upon receipt of the intervention itself, without contamination from the experience of taking the quantitative assessments associated STOP when delivered as part of the main trial.

METHODS – Participants, interventions, and outcomes

Study setting {9}

This study will be conducted across two sites. The lead site is the Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London. The second site is the Department of Psychology, University of Bath.

Eligibility criteria {10}

Eligible participants will be outpatients with distressing paranoid symptoms, irrespective of primary diagnosis. This is appropriate because STOP is designed to target distress caused by paranoid symptoms rather than a specific diagnosis. SCID-V Research version (SCID-5-RV) and PANSS clinical interviews will be conducted during baseline assessment to establish diagnoses and comorbidity. Potential participants will be screened and selected for invitation to participate by researchers according to the inclusion and exclusion criteria.

The **inclusion criteria** are:

1. **Any clinically significant persecutory or paranoid symptoms, present for at least the preceding month.** This will be operationalised as a **score of 3 ("mild") or more on item 6 of the Positive and Negative Syndrome Scale (PANSS, [25]).** Item 6 measures "suspiciousness/persecution" defined as 'unrealistic or exaggerated ideas of persecution, as reflected in guardedness, a distrustful attitude, suspicious hyper-vigilance, or frank delusions that others mean one harm'. A clinician rating is made on the basis of the patient's thought content and its influence on behaviour as expressed at interview in response to 7 prompt questions. Scores range from 1 (absent) to 7 (extreme). Paranoid symptoms will be assessed remotely or in-person by a trained researcher.

Rationale: This inclusion criterion is deliberately transdiagnostic and reflects a functional approach to treatment, a desire to avoid stigmatising labels and reflects preferences expressed during service user feedback on study design. A lower threshold of 3 is selected because 2 does not meet clinical levels (defined as 'Minimal - questionable pathology; may be at the upper extreme of normal limits'), whereas 3 represents the start of clinical levels of symptomatology (defined as 'Mild - presents a guarded or even openly distrustful attitude, but thoughts interactions and behaviour are minimally affected').

2. **Displaying an interpretation bias ≥ -2 on the 8-item screening version of the Similarity Ratings Task (SRT).** Potential participants will be asked to complete a bias screening task online (for those without internet access, a paper copy will be sent to their

home, or the task can be administered by phone). The SRT screening task is a shortened (8-item) version of the Similarity Rating Task (SRT) [26, 27]. In the screening SRT, sentences reflecting non-paranoid (i.e., benign) and paranoid interpretations are denoted TNpar and TPar, respectively, and rated by participants on a 1-4 Likert scale (higher score = stronger endorsement). An interpretation bias screening score is calculated: TPar minus TNPar. The screening score has a range of +3 (paranoid bias) to −3 (non-paranoid bias). A bias score of -2 or greater (i.e. -2 up to +3) is necessary to be selected for the study.

Rationale: This inclusion criterion is adopted in order to screen out any patients already displaying a strong preexisting non-paranoid bias. Such patients would be unlikely to benefit from an intervention whose main mechanism of action is to create a *more* non-paranoid bias, since they would already be at or near ceiling and are not showing the maladaptive thinking pattern that the intervention targets.

3. **If on psychotropic medication, then stable on that medication for at least the last 3 months and expected to be so for the study duration**. Psychotropic medication has been shown to influence the kind of cognitive biases targeted by STOP and as such the rationale for adopting this inclusion criterion is the same as described below, under exclusion criterion 5.

4. Age 18 years or over.

5. Capacity to consent; if in doubt assessed using the Capacity Assessment Tool.

The **exclusion criteria** are:

1. Severe cognitive impairment;

2. Illiteracy or inability to understand written and spoken English for any other reason;

3. Major current physical illness (e.g., cancer, heart disease, stroke, dementia);

4. Major substance or alcohol misuse, assessed by the SCID-V screen;

5. Currently receiving, or soon due to receive, a psychological intervention* targeting the same psychological mechanism as STOP (i.e. paranoid belief change) or having done so in the last 3 months;

6. Currently taking part in any other interventional research study;

7. Scoring 7 (defined as ‘Extreme - a network of systematised persecutory delusions denotes the patient’s thinking, social relations and behaviour’) on item 6 of the PANSS.

8. At high risk of suicide as indicated by the Columbia Suicide Severity Rating Scale (C-SSRS) - Screen Version [78,79]. This instrument uses a series of standardised questions covering ideation, intent, plans and behaviour over the last 1-3 months. Those falling into the high-risk category (i.e., expressing intent -with or without a plan- or reporting a suicidal behaviour) will be excluded and their referring clinician alerted.

Exclusion criteria 7 and 8 are introduced in order to limit the clinical risk associated with vulnerable patients taking part in a clinical trial of a 'Software as a Medical Device' (SaMD) product.

*Psychological intervention is defined as five or more consecutive sessions with a trained therapist, either in an individual or group setting, for the purpose of alleviating symptoms arising from paranoia or related issues. Receipt of intervention for any other purpose (e.g., occupational or art therapy; counselling) will be assessed on a case-by-case basis, against the principle that eligibility requires any cognitive mechanisms targeted by therapy to be independent of those targeted by STOP.

Rationale: Many psychotherapies (e.g. cognitive behavioural therapy) include a focus on the same psychological mechanism as STOP. If ignored, this could seriously confound the interpretation of the study findings, as these participants would be receiving an unquantified enhancement to the treatment dose for the target mechanism. Furthermore, unknown interactive effects could compromise either or both treatments making it, arguably, unethical to combine the two. Only where clinical judgement confirmed that there was no risk of these problems was a patient permitted to join the trial at the same time as receiving another psychological intervention.

Who will take informed consent?
Capacity to consent will be assumed in accordance with the principles of the Mental Capacity Act 2005 [28]. Where there is doubt, capacity will be assessed by researchers during screening using a capacity assessment tool that is based on the Mental Capacity Act 2005 (which assumes capacity by default unless there is reason for doubt). Participants will be deemed to lack capacity in the unlikely event that they are unable to understand and retain information provided to them about the study and their involvement, and/or they cannot communicate their decision, all of which is determined by the questions contained within the standard written protocol.

The study researchers at each site will provide potential participants with information about the STOP study by reviewing the Participant Information Sheet with them and answering any questions they have. They will then be offered at least 24 hours to consider all the information provided before written consent is obtained.

Written informed consent will be documented on the consent form, with both a participant and researcher signature. The original copy of the consent form will be retained in the Trial Master File. If consent is obtained remotely (i.e. by phone), the researcher will sign the consent form and make a note that verbal informed consent was obtained from the participant.

Additional consent provisions for collection and use of participant data
Data Deposit
We will include an explicit item in our informed consent procedures explaining the process for fully anonymised data sharing via the national repository, UK Data Archive and data files will be deposited at the ISRCTN registry (https://www.biomedcentral.com/p/isrctn). Participants who do not wish their anonymised data to be deposited can avoid checking this box and/or strike out this item.

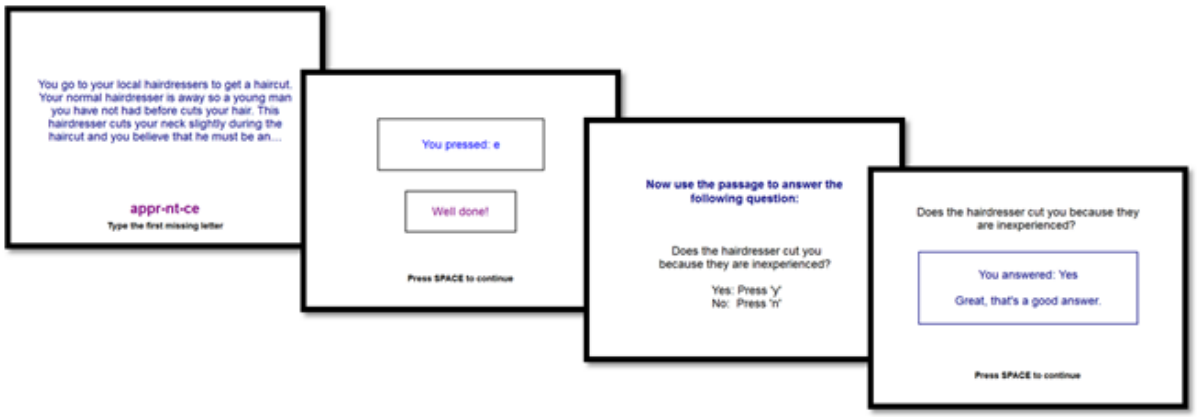
Interventions

Explanation for the choice of comparators
The study uses an active control condition as this directly evaluates the effect of specifically manipulating the hypothesised mechanism of action (bias reduction) and controls for any non-specific therapy effects (e.g. contact with researchers, interaction with digital equipment, trial participation etc.) in a manner not possible with a waiting list control.

Intervention description

STOP Intervention
Participants randomised to this condition will receive 40 training items per session (approximately 40min duration) on the STOP mobile App. Participants will read text inviting paranoid interpretations, then complete missing words and answer questions in a way that encourages alternative, adaptive beliefs about themselves and others (see example given in Figure 2).

Figure 2. Training item example



In addition to the existing 240 training items (40 training items across six sessions) from the CBM-pa feasibility study, we will test a newly created additional set of 240 training items for the additional six sessions (i.e., sessions 7-12 for the 12-session arm). We maximise clinical relevance by gathering written examples from patients of everyday situations that could trigger paranoia drawn from their personal experiences. These are used as the base content and further developed by the research team into the format required for use as a training item. In this way item content, including scenarios, questions and responses was derived from real life examples generated by 18 service users with personal experience of clinical paranoia. The academic team revised the questions and answers into standard format shown in Figure 2 (3-line stories with a final word completion resolving ambiguity positively, followed by a one sentence question to reinforce the positive resolution).”

The order of intervention sessions is informed by the paranoia hierarchy [70]. Clinician-delivered exposure therapy involves gradually exposing clients to increasingly triggering situations meaning early sessions are less anxiety-provoking and more tolerable for patients. This allows habituation of the emotional response to occur gradually as the patient progresses through to more challenging therapeutic content, and is usually more acceptable to patients, reducing the risk of dropout. We sought to emulate this process in STOP, following previous

basic research suggesting that a similar approach can be usefully applied to CBM therapies [71]. Early sessions of STOP include less severe items, with a graded progression toward more potentially threatening items in later sessions.

To achieve this the following methodology was used. Each item was rated for severity on a 6 -point Likert scale by 4 clinical psychologists with at least one year of experience with paranoia. Items were then rank ordered by mean severity rating and divided into separate sessions of 40 items per session. In order to understand, and as far as possible balance, the content coverage within individual sessions, items were categorized by researchers into the follow topic areas:

1. Social/interpersonal threat
2. Physical harm
3. Threat of persecution/spying
4. Delusions of reference/magical thinking
5. Medical/paramedical/health care threat
6. General suspiciousness/distrust

Where possible topics 1-6 were evenly distributed across sessions. However, some topics were inherently more severe and therefore clustered towards the later sessions (e.g. categories 4, 5) whereas others were inherently milder and therefore skewed towards earlier sessions (e.g. 1).

The item sets were then subject to three 2hr video-conferencing meetings conducted in collaboration with the McPin Charity. 4-6 users attended each meeting to review scenarios, including each question and answer. One week before each meeting, users were sent a batch of 80 items to review in a word document with a comment box that allowed user time to preview the items, make free text comments and flag any items to be discussed during the meeting. During each meeting, we reviewed the content item by item by presenting it in a word document using screen sharing until all 80 items had been reviewed and all comments considered. Items (text, word completion and question/answer) were collaboratively discussed as necessary; users could either make verbal suggestions or type their comments in the chat function. With participants' consent, the meetings were recorded and the automated transcript used to inform subsequent changes to items, after which it was deleted. To provide further opportunity for users to comment in light of the live discussions, after each meeting, we invited users to provide any further feedback and suggestions in a word document.

Taking the above iterative feedback into consideration, further replacements/revisions were made by the academic team. 18 scenarios including questions and answers required major revisions; 4-6 service users were again invited to review these revised items to make further comments and agree the final set.

Finally, clinician administered cognitive therapies traditionally employ an increasing 'drill-down' from more surface-level automatic thoughts, into more rigid rules and assumptions, and then finally towards deeper core beliefs. In order to reflect this in STOP we selected specific verbs for the final sentence of each passage. In sessions 1-4 the following verbs were used to reflect automatic thoughts: 'think', 'imagine', or 'sense' (e.g., you *think* that this means…). In sessions 5-8 the following verbs were used to capture underlying assumptions: 'assume', 'presume', 'suppose' (e.g., you *assume* that this means …). In sessions 9-12 core beliefs were reflected using the verbs 'believe', 'are sure', 'know' (e.g., you *believe* that this means …). Appropriate verbs to use to capture each level were decided by consensus of 3

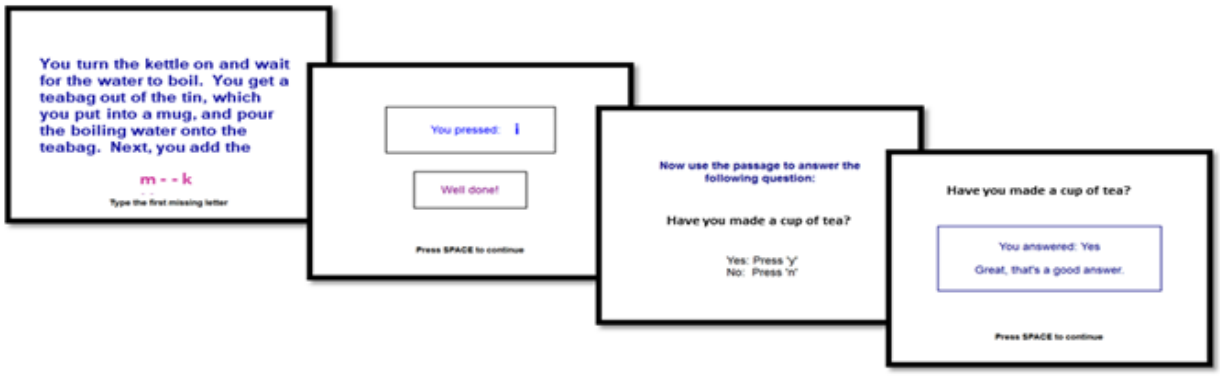
members of the research team with experience in paranoia, cognitive therapies and CBM methods.

In the six-session arm of STOP users receive only the first 6 sessions of the full 12 session therapy. The content of the first 6 sessions are therefore less severe in terms of quantitative paranoia ratings and use only the earlier verbs in the drill down sequence.

Text-reading control

Participants randomised to the active control condition will also receive 40 control items per session (approximately 40min duration) on the STOP mobile App. The experience will be identical to the intervention condition except the item content that participants see omits the active ingredient: resolution of an emotionally ambiguous situation in a benign/non-paranoid manner. Instead, control participants read and respond to factual material or mundane everyday experiences (see example given in Figure 3).

Figure 3. Control item example.



Assessment of Contamination Risk

Contamination is the process whereby an intervention intended for members of one arm (the experimental intervention or treatment arm) is received by members of another (the control). Contamination risk is often a concern in trials of psychological and/or behavioural interventions where masking to trial arm allocation may be more difficult to achieve. We assessed potential contamination risk at the design stage using the framework developed by Jacobsen and Wood (2020) [29]. This framework involves assessing potential contamination risk via the three possible sources of contamination: A) participant in control arm accesses treatment in the intervention arm themselves; B) participant in intervention arm passes treatment on to participant in control arm; C) therapist providing treatment in intervention arm passes treatment on to participant in control arm. We assessed the risk of contamination as being low across all sources given that: i) the app is not freely available outside of the clinical trial; ii) personalised log-ins prevent sharing of the app between trial participants in different arms of the trial, iii) the intervention is a digital intervention rather than being delivered by a therapist.

Criteria for discontinuing or modifying allocated interventions{11b}
The trial arm interventions cannot be modified for individual participants. Participants choosing to take part will follow the procedures as predefined for their allocated arm.

a) Deterioration of mood. State mood will be measured before and after each session using the visual analogue scales anxious, sad, paranoid, friendly [30]. State mood normally fluctuates widely on a day to day and moment to moment basis, however a consistent pattern over time can indicate more enduring effects that should be investigated further. Therefore, an automatic alert will be sent to researchers if a participants' mood worsens across the session on 3 consecutive sessions. Following an alert, a researcher will contact the participant to make a further assessment and escalate within the research team if appropriate, and, if necessary, procedures for notifying an adverse event or SAE will be triggered.

b) Adverse events. At the trial level we will take all appropriate steps during the conduct of the trial for ensuring participant safety, in all arms of the trial. Any arising concerns for participant well-being, events or incidents will be reported through standardised Adverse Events (AEs) and Serious Adverse Events (SAEs) reporting procedures created for this study based upon templates provided by the King's CTU and other similar trials. These procedures involve immediate notification to the DMC/TSC Chairs of any SAEs, in addition to regular scrutiny of AE rates by arm at the committees' scheduled meetings. If deemed sufficiently serious and caused by the trial or related procedures by either the study team or the DMC/TSC, this could lead to study termination at any time. Experience with this population in the feasibility study using a 6-session desktop version of the mobile app suggests that is extremely unlikely.

c) Unmet Milestones. The trial may be prematurely discontinued by the MRC should key Milestones remain unmet or based on new safety information or for other reasons. The DMC/TSC can also independently recommend discontinuing the study. If the study is prematurely discontinued, active participants will be informed and withdrawn and no further participant data will be collected. Existing data will be retained in line with the data management plan.

Withdrawal of Participants

Participants have the right to withdraw from the study at any time for any reason. It will be made clear to each participant at the time of informed consent that they may withdraw from the trial at any time without having to give a reason and without this impacting on their usual clinical care in any way.

The investigator also has the right to withdraw patients from the study intervention in the event of inter-current illness, AEs, SAE's, protocol violations, administrative reasons or other reasons. Participants may also be withdrawn from the trial in any of the preceding circumstances (a-c) if judged appropriate by the participant themselves, the Trial Steering Committee or the Chief Investigator.

It is understood by all concerned that an excessive rate of withdrawals can render the study uninterpretable; therefore, unnecessary withdrawal of patients will be avoided. Should a patient decide to withdraw from the study, all efforts will be made to report the reason for withdrawal as thoroughly as possible. Should a patient withdraw, or be withdrawn, from study drug intervention only, efforts will be made to continue to obtain follow-up data, with the permission of the patient.

Device withdrawal procedures

Provision for product recall, for example in the event of a serious adverse event, serious technical problem or for any other reason, exists at both the individual (single user) and aggregate (i.e. all users) level.

Individual level. Individual users can be prevented from using the app if necessary (for example should individual level clinical safety concerns arise) by the developer (Avegen) disabling that user's login credentials, upon instruction from the research team. Should this be necessary, a member of the research team will contact the user (and, if appropriate, their responsible clinician) to manage the situation and make the user aware that their app usage has been disabled. Users will see the following recall message: "*You can no longer use STOP as your registration has been withdrawn for safety reasons. You will shortly be contacted by a member of the research team who will explain and offer further support if required. In the meantime, if you require more urgent assistance, please contact the study helpline at 07897408675 or clinical support email at stopclinicalsupport@kcl.ac.uk*."

Cohort level. A full product recall can be implemented if necessary (for example, should safety concerns arise that could affect all users) using a 'recall switch' feature within the app.
In the unlikely event that a full product recall is required during the investigation, the following will occur:

1. STOP users will see the following recall message: "*You can no longer use STOP as the product has been withdrawn. You will shortly be contacted by a member of the research team who will explain and offer further support if required. In the meantime, if you require more urgent assistance, please contact the study helpline at 07897408675 or clinical support email at stopclinicalsupport@kcl.ac.uk*".
2. Researchers will systematically contact every individual currently enrolled in the trial to explain the reasons for the product recall and offer additional support as necessary.
3. In the event of a SAE leading to a complete product recall the responsible clinician for every currently enrolled patient will be informed of the event, and adverse event reporting procedures outlined in the SOP will be followed to inform relevant authorities, including MHRA and the Ethics committee.

To permit a full product recall, as well as for routine troubleshooting, and to support future product Post Market Surveillance and Product Improvement, the following data related to every user's device is automatically collected:

a. The app version
b. Platform OS
c. Model name
d. Last authentication date
e. Last login by the user

This is explained to users in the STOP app Privacy Notice and Participant Information Sheet.

Strategies to improve adherence to interventions {11c}

Adherence is a known challenge faced by online interventions. To mitigate this, STOP will include proven gamification techniques including progress tracking and reward systems.
During trial, the following will reduce drop-out:

1. When registering patients into the trial, researchers and participants will together schedule dates for completing each session using the app's calendar. A (blinded) administrator portal will track participants' progress. Missed or incomplete sessions or assessments will trigger auto-alerts to the team for investigation and/or additional support.

2. Participants (intervention and control) will receive a weekly or fortnightly phone call from researchers, following a pre-prepared script to control bias.
3. STOP will include **graphics enhancements**. In the CBM-pa feasibility trial, end-users recommended visually enriching trials in addition to text passages to make scenarios more engaging and realistic. Further, researchers have shown that the effectiveness of the clinical intervention is positively correlated with the degree of participants' engagement [31]. Both STOP and the text reading control condition will therefore include audio visual enhancements (e.g. dynamic and static images and sounds) to increase participants interest and engagement with the text they are reading. A written report from a graphics strategy workshop with potential end users will be used to define the most appropriate enhancements to use.

Adherence to the intervention is defined as outlined in the table below.

Table 1. Definitions of Adherence

		Acceptable Values	**Target Values**
Individual sessions	To 'complete' an individual session participant must….	enter responses on at least 20 out of 40 trials (50%)	enter responses on at least 30 out of 40 trials (75%)
Intervention overall	To 'complete' the intervention as a whole participants must….	Complete ¼ of all sessions scheduled to date	Complete ½ of all sessions scheduled to date
Number of participants	**Proportion of sample to be adherent in order to meet funder milestones**	**50%**	**75%**

The following boundaries will apply to the delivery of interventions and assessments, all of which are incorporated within the programmed functionality of the app and are therefore implemented automatically:

Interventions

- The intervention will comprise 12 predefined, consecutive calendar weeks for each participant, starting on the first Saturday after date of randomisation.
- Each calendar week runs from 00.00 Saturday – 23.59 Friday. Saturday is chosen as the start of the week to allow maximum time for users to miss and reschedule their sessions.

- Participants and researchers will schedule the first session together, which then prepopulates the remaining sessions, all ideally 7 days apart.
- Researchers will ensure that there are at least two days between randomisation and the first scheduled session (a technical requirement of the app software system).
- Subsequent sessions can be individually adjusted (e.g. to accommodate other commitments) as needed within the following allowable parameters:
    - Each session must occur at sometime within its 7 day allocated calendar week (this prevents problematic variance in intensity of treatment if using a 7 days +/- 2' condition where one ppt can always go for 5 days apart – giving a 60 day tx window - and another 9 days apart – giving a 108 day tx window).
    - Each session must be at least 2 full days apart (thus Friday of week 1 can be followed by either Monday of week 2, or the last day of week 2, Sunday. Session 3 would then have to fall, in the first case, between Monday and Sunday of week 3 and in the second case between Wednesday and Sunday of wk 3. Sessions can therefore be between 2 and 12 days apart)

- Participants are encouraged to schedule all sessions early in the calendar week (i.e. close to Saturday) to allow maximum time for rescheduling if needed, thereby reducing the risk of missed sessions.
- Participants are asked to plan to complete each session in a 'single sitting', and have the option to take short comfort breaks during the session after every 10 items. If unavoidably interrupted for longer periods during a session they should complete within the same day. Sessions not completed the same day will trigger an alert to researchers the next day who can follow up case by case.
- Sessions remain open for completion up until 2 days prior to the next scheduled session (i.e. consistent with the minimum allowable time between consecutive sessions)

Assessments

- Follow up assessment dates are calculated as 6, 12, 18, 24 weeks post randomisation date.
- A window of +/- 3 days either side is allowed, with the exception of the 6 wk assessment, which is automatically scheduled at the mid-point between session 6 and session 7.
- Researcher contact is required at each assessment to administer PANSS 6 and maintain contact.
- All participants receive all 12 interim assessments (those randomised to the 6-session arm will receive only assessments in weeks 7-12). This serves 2 purposes: a) provides control data for the secondary dose response analysis and b) prevents unblinding that could result from scheduling a different number of sessions.

Relevant concomitant care permitted or prohibited during the trial {11d}

Care permitted: all three conditions will be conducted in addition to TAU, which will be in the form of individualised combinations of medication and care coordination.

Care prohibited: None. If a patient starts to receive a psychological intervention targeting the same psychological mechanisms as STOP (expected receipt of such treatment during the trial is an exclusion criterion, see above) during their participation in the trial, then they will continue to participate in the trial. This will be captured by the routine TAU monitoring carried out at each main outcome assessment and cases where this occurs will be flagged in the final dataset for analysis with and without these cases.

Patients are not expected to have an inpatient admission, since only those in a stabilised condition are suitable for the trial (see inclusion/exclusion criteria). However, should an inpatient admission occur while a patient is participating then active participation may continue unless practical constraints (e.g. no access to a phone) or clinical advice from the patient's care team prevent it. An inactive participant (e.g. missing sessions due to an inpatient admission) will remain registered on the trial unless they choose to withdraw and may resume active participation in the trial as and when their circumstance permit, should they wish to do so.

Provisions for post-trial care {30}
TAU will continue unchanged upon participants exit from the trial.
In the event that participants are harmed during the research they may have grounds for legal action for compensation against King's College London, Bath University and/or their local NHS Trust, but they may have to pay their legal costs. The normal National Health Service complaints mechanisms will be available to participants (if appropriate). King's College London has obtained insurance which provides no-fault compensation (i.e., for non-negligent harm), and participants may be entitled to make a claim for this.

Outcomes {12}

The full assessment schedule is shown in Figure 4.

Figure 4. Assessment Schedule

	Timeline	Screening	Baseline 0 weeks	Session 1-6	Follow-up 6 weeks (this is end of treatment for the 6-session arm, mid treatment for the other arms)	Session 7-12 (The 6-session arm does not receive these treatment sessions, but does receive assessments)	Follow-up 12 weeks (this is 3 mth post randomisation FU for 6 session arm, end of treatment for others)	Follow ups 18 weeks and 24 weeks
1	***Socio-demographics***							
2	Inclusion/Exclusion criteria checklist							
3	Capacity Assessment Tool							
4	Socio-demographic data							
5	Participant Consent Form							
6	Participant Information Sheet							
7	Credibility/Expectancy Questionnaire							
8	Wechsler Test of Adult Reading (WTAR)							
9	Persuadability subscale (PER)							
10	***Target mechanism (bias)***							
11	Similarity Ratings Task (SRT)		C/bal					C/bal
12	Scrambled Sentences Task (SST)		C/bal					C/bal
13	***Clinical Symptoms***							
14	Paranoia Scale*		C/bal					C/bal
15	Green Paranoid Thought Scale (GPTS) short version (18 items)		C/bal					C/bal
16	PANSS- item 6 only							
17	SCID-5 Diagnostic interview							
18	HADS							
19	Paranoia Worries Questionnaire							
20	***State Mood change***							
21	VAS (anxious, sad, paranoid, friendly)			Pre-post intervention		Pre-post intervention		
22	***Recovery***							
23	WEMWBMS							
24	QPR							
25	ReQoL							
26	User Satisfaction (Ben-Zeev et al 2014)							FU 24w only
27	***Miscellaneous***							
28	Treatment As Usual							
29	Self-reported Treatment Allocation							FU 24w only
30	Negative Effects Questionnaire							

Notes

* indicates Primary Outcome
Grey cells indicate measures subject to counterbalancing (C/bal) for 1. task order (indicated by vertical arrows) 2. item content across multiple uses of the same task (indicated by horizontal arrows). Full counterbalancing schedules available upon request.
Green cells indicate baseline assessment measures.
Pink cells indicate follow up assessment measures.
Blue cells indicate other assessments (screening, pre-post mood checks, interim)

PANSS: The Positive and Negative Syndrome Scale; HADS: Hospital Anxiety and Depression Scale; VAS: Visual Analogue Scale; WEMWBMS: Warwick-Edinburgh Mental Wellbeing Scale; QPR: The Questionnaire about the Process of Recovery; ReQoL: Recovering Quality of Life

Primary outcome measure
The primary outcome is mean paranoid symptoms measured using the Paranoia Scale (PS) [24] at 24 weeks post-randomisation. The PS is a 20-item self-report measure designed to assess paranoid ideation in response to everyday events and situations. Items on the PS include, "someone has it in for me" and "it is safer to trust no one". Participants rate their agreement with each item using a 5-point scale (1=*not at all applicable to me* to 5 = e*xtremely applicable to me*). Individual scores are summed, ranging from 20–100 with higher scores indicating greater endorsement of paranoid ideation. Mean change in score from baseline is aggregated and recorded for analyses. The PS has been used widely within clinical research with paranoia and has good psychometric properties with internal consistency, r=0.84, and test-retest reliability, r=.70 [24].

Secondary outcome measures
1. Paranoid symptoms measured using the Paranoia Scale at baseline, 6, 12, 18 and 24 weeks post-randomisation
2. Paranoid beliefs (the targeted mechanism of interpretation bias) are measured at baseline, 6, 12, 18 and 24 weeks post-randomisation using:

- **Similarity Rating Test** (SRT) [26, 27]. In the SRT, sentences reflecting non-paranoid (i.e., benign) and paranoid interpretations of a previous 3-sentence ambiguous passage are denoted TNpar and TPar, respectively. Participants rate 'how similar in meaning is this sentence to the passage you saw earlier?' use a Likert scale, 1 (= very different in meaning) to 4 (= very similar in meaning). A higher score therefore reflects stronger endorsement for how similar a sentence is to the previous passage. Mean ratings are calculated for each type of sentence separately and then an interpretation bias score is calculated: mean [TPar] minus mean [TNPar]. Bias scores therefore range from +3 (strong paranoid bias) to -3 (strong non-paranoid bias). The SRT bias score has acceptable reliability (Cronbach's $\alpha = 0.63$) [32].
- **Scrambled Sentences Task** (SST) [33, 34] is a robust measure of interpretation bias that requires participants to re-order five out of six words to create a grammatically correct sentence (e.g. 'are me hostile strangers to friendly'). Participants have 5 minutes to complete as many of the 15 sentences as possible. Before unscrambling the sentences, participants are asked to memorise a six-digit number (e.g. 615239) for a later recall test and keep it in mind while completing the task. Each word string can be unscrambled into sentences with either paranoid or non-paranoid meaning. Paranoid and non-paranoid interpretation bias scores are calculated as a percentage of either paranoid or non-paranoid (respectively) interpretations made by recording: the total number of grammatically correct sentences created with a paranoid (or non-paranoid) meaning (A) and the total number of responses within the available time (B), then calculating: [A/B] × 100. Bias scores therefore range from 0 to 100 % with higher scores reflecting larger bias favouring the specified direction of interpretation (paranoid or non-paranoid). Full scoring instructions used are available here. SST bias scores have good reliability (Cronbach's $\alpha = 0.73$) [35].

3. Clinical symptoms are measured at baseline, 6, 12, 18 and 24 weeks post-randomisation using:

- Green Paranoid Thoughts Scale Revised (R-GPTS) [36]. The scale is an 18-item trait measure of paranoia, with two subscales assessing ideas of reference and ideas of persecution. Participants rate their agreement with each item using a 5-point scale.

Individual scores are summed to give a total score per participant, ranging from 0–72 (ideas of reference scale: 0–32; ideas of persecution scale: 0–40) with higher scores indicating greater severity of paranoid thinking. The R-GPTS retains good reliability across the spectrum of paranoia severity, r = .97 [36].

- Positive and Negative Symptom Schedule (PANSS) - item 6 only [37]. Item 6 of the positive symptom subscale measures suspiciousness and persecutory ideations defined as 'unrealistic or exaggerated ideas of persecution, as reflected in guardedness, a distrustful attitude, suspicious hyper-vigilance, or frank delusions that others mean one harm'. A clinician rating is made on the basis of the patient's thought content and its influence on behaviour as expressed at interview in response to 7 prompt questions. Scores range from 1 (absent) to 6 (severe) with higher scores reflecting higher paranoia. The scale has good psychometric properties for measuring symptomatology of paranoia, with r=.72, internal consistency across positive scale [37] and inter-rater reliability ranging from r=.83–.87.
- Hospital Anxiety and Depression Scale (HADS) [38]. The scale is a 14-item measure of anxiety and depression. Participants rate their agreement with each item using a 4-point scale. Individual scores for each scale (depression; anxiety) are summed, ranging from 0–21 with higher scores indicating higher level of anxiety and depression. The HADS has been used widely within clinical research across different populations and has good psychometric properties with internal consistency, r > .80 [39, 40].
- Paranoia Worries Questionnaire (PWQ) [41]. The PWQ is a 5-item measure designed to assess the severity of paranoia worry. Participants rate their agreement with each item using a 5-point scale. Individual scores are summed, ranging from 0–20 with higher scores indicating more severe levels of paranoia worry. The PWQ has good psychometric properties, with internal consistency, r > 0.95, and test-retest reliability, r=.96, over a 1-week period [41].

4. Recovery is measured at baseline, 6, 12, 18 and 24 weeks post-randomisation using:

- The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) [42]. The WEMWBS is a 14-item measure of mental wellbeing in the general population, assessing positive affect and functioning and satisfying interpersonal relationships. Participants rate their agreement with each item using a 5-point scale. Individual scores are summed, ranging from 14–70 with higher scores indicating a higher level of mental wellbeing. The WEMWBS has good psychometric properties for evaluating mental wellbeing (internal consistency, r = 0.91; test-retest reliability at one week, r = .83 [42].

- Questionnaire about the Process of Recovery (QPR) [43]. The QPR version used here is a 15-item measure of recovery from psychosis, including general psychological wellbeing, quality of life, and empowerment, allocated to two subscales (intrapersonal and interpersonal). Participants rate their agreement with each item using a 4-point scale. Individual scores are summed, ranging from 0–60 with higher scores indicating better recovery from psychosis. The QPR has been widely used in both clinical practice and research and has good psychometric properties for evaluating recovery from psychosis (internal consistency: intrapersonal r = .94; interpersonal r = .77; 2-week test-retest reliability: intrapersonal r = .87; interpersonal r = .77 [44].

- Recovering Quality of Life (ReQoL) questionnaire [45]. The ReQoL version used here is a 10-item measure designed to assess the outcome of mental health services. Participants rate their agreement with each item using a 5-point scale. Individual scores are summed, ranging from 0–40 with higher scores indicating highest quality of life. The ReQoL has good psychometric properties for evaluating quality of life, with a reported internal consistency, r = .92 [45].

User satisfaction with the mobile app is measured with the Ben-Zeev et al. [46] instrument given once only at the 24-week follow up.

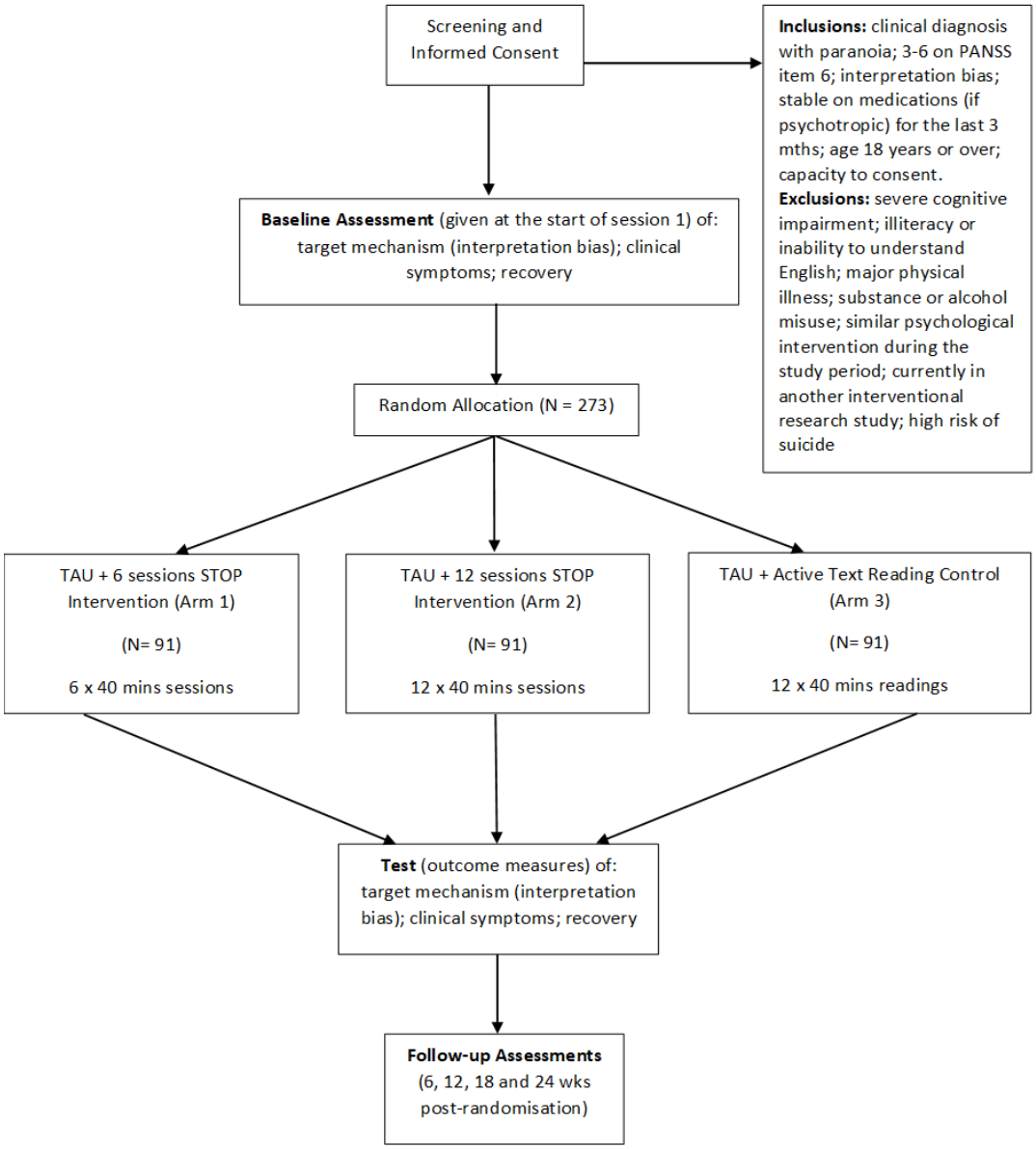
Adverse and unwanted effects of the intervention will be assessed quantitatively using the Negative Effects Questionnaire (NEQ; [80]). This measure was chosen because it is a validated, feasible/low burden (20 items), self-report measure that can be completed by patients in English and addresses stigma and deterioration of symptoms that are relevant to digital interventions and it was developed from the results of a consensus statement [75]. It will be given to all participants at 12 week follow up to compare the negative effects of STOP in the treatment arms versus the control arm.

State mood change will be measured before and after each session using Visual Analog Scales (VAS: anxious, sad, paranoid, friendly) [47]. VASs for mood are psychometric response scales used to measure subjective mood using visual representations (e.g., faces representing different mood states). These will help us track, weekly, any abrupt (and progressive) changes in mood and paranoid symptoms, thereby monitoring different psychological aspects of each participant over time and assessing the safety of the STOP mobile App.

Participant timeline

Participants will be randomised into one of the three study arms. The detailed timeline of assessments is shown in the Assessment Schedule, above. The participant pathway through the trial is shown in the Trial Flow Diagram below.

Figure 5. Trial Flow Diagram



The first contact with participants will be at screening, where we check for participants' eligibility by assessing the inclusion and exclusion criteria described above and complete a capacity assessment, where this is in doubt (see 'Who will take informed consent' for further details).

Eligible participants interested to take part will be provided with the ethically-approved final version of the study information sheet, with at least 24 hours to consider taking part.

Those subsequently electing to take part will meet remotely or in person with a trained researcher who will administer informed consent and baseline assessment.

Randomisation

Following baseline assessment, researchers will randomise the participant using the services of the King's Clinical Trials Unit (see 'assignment of interventions', below). Patients will be told: "You will be randomised to one of three different procedures of varying lengths, some of which we think may help reduce symptoms of paranoia. 'Randomised' means being put into a group completely by chance, it is a bit like flipping a coin. You will have an equal chance of being put in any of the three groups." Participants can remain blind to treatment arm because the active control condition is procedurally identical to the treatment (compare Figure 2 and 3) and all participants will receive all assessments (those randomised to the 6-session arm will continue to receive weekly assessments for the full 12 weeks). Further details of randomisation procedures are given under section Methods- Assignment of Interventions, below.

Onboarding

The STOP intervention or text control is delivered on mobile phones, set up by the researchers, and thereafter is a self-directed, automated package, self-administered by the patient.
Participants will be given the choice to use either their own phone or a study phone and their choice and reasons for it will be recorded. Researchers will register participants and their phones into the trial using a two-factor authentication process administered via an online dashboard (HealthMachine) which is integrated with the STOP app.

Researcher contact points

Researchers will have direct contact (either face to face or virtual) with participants to carry out the following tasks: screening, baseline assessment, onboarding to the STOP app, follow up (but not interim) assessments.

Check-in phone calls. In addition, consistent with feasibility study feedback, participants (intervention and control) will receive a fortnightly phone call from researchers, following a pre-prepared script to limit bias. Participants who score >58 (2.5 standard deviations above the general population mean) on the Persuadability subscale of the Multidimensional Iowa Suggestibility Scale (MISS; [77]) will be offered additional support in the form of weekly (rather than fortnightly) check-in phone calls. Key objectives of this contact point will be: i) to check participants well-being ii) record any adverse events ii) promote engagement in the trial. Missed sessions or assessments, or those not fully completed the same day, will trigger an alert to researchers the next day who will make contact with participants to check well-being and encourage adherence (see Table 4, below, for schedule of alerts and notifications)
Three consecutive sessions (in any arm) showing worsening state mood on one or more visual analogue scale will trigger a researcher alert to check participant well-being.

Sample size {14}

With alpha = .05 and 80% power, recruiting a sample of 77 per group would permit detection of a clinically useful drop of 8 points on the Paranoia Scale compared to TAU assuming a standard deviation of 17.5 (based on feasibility data). This magnitude drop would be a 15% reduction in the average score of paranoid patients (52: [48]) and represents an effect size of d=0.46 (small to medium). Assuming a conservative estimate of 15% drop-out (from feasibility data) a total sample size of 273 is required when using an independent t-test.

Recruitment

Recruitment will be carried out by two researchers (one clinical researcher plus one research assistant) at each site, liaising with secondary care services to receive appropriate referrals by

clinicians of patients likely to meet the study's inclusion and exclusion criteria. A Patient and Public Involvement (PPI) staff member from The McPin Foundation will also provide recruitment through service user networks, which was very successful in the feasibility study.

The recruitment target for The McPin Foundation is 50 eligible participants based on contributions during the feasibility study. Allowing for this recruitment sample, the sites' recruitment target is 223 (273-50) with a target referral rate of 17 patients per month. This should deliver 10 eligible study participants per month according to the feasibility data, which showed that 60% of referrals were eligible to take part. Over the 23 months scheduled for the trial this would elicit a sample of 230. At site level this requires a minimum of 2 referrals/1 eligible participant per week per site.

Table 2. Recruitment Targets

	Overall	Per Site
Monthly		
Referrals	17	9
Eligible participants	10	5
Weekly		
Referrals	6	3
Eligible participants	4	2

Participants will be recruited from a range of local sources in both London and Bath, including:

- Local/regional NHS Trusts – In London, we will recruit from three local trusts (South London and Maudsley (SLaM), Oxleas, and Kent and Medway). In Bath, two local trusts (Avon and Wiltshire, and Gloucestershire) will participate
- Patient recruitment registers – Consent 4 Contact (C4C); Cognition, Schizophrenia and Imaging Laboratory (CSI Lab); Psychological Interventions Clinic for Outpatients with Psychosis (PICuP); and the OASIS early intervention clinic
- Service user/self-help networks
- Public advertisements
- Trial website
- Word of mouth from existing participants
- Ongoing studies with similar inclusion criteria
- Therapy waiting lists

METHODS – Assignment of interventions

Allocation:
Sequence generation

Randomisation is at the patient level and is performed by software development company Avegen, with advice and input from King's Clinical Trials Unit (KCTU). The randomisation procedure will be completed in the STOP app. There will be no adjustments permitted to randomisation once completed. Researchers will remain blind to allocated condition.

Researchers will be registered on the KCTU system and assigned usernames and passwords. They will log into the KCTU MACRO database system in order to generate a unique PIN for each participant prior to screening (see implementation, below).

At screening, researchers will enter that PIN into Health Machine (HM; the researcher platform for administering the STOP app) which will register the participant for eligibility assessment for the STOP trial. Screening and, if eligible, subsequent baseline assessment will be completed using a combination of Health Machine,the STOP app and Client Report Forms (CRFs).

Randomisation will take place at the end of baseline assessment using a checkbox within the HM platform to provide an exportable timestamp. At this point the PIN will be randomly allocated to one of the three study arms automatically by the HM/ STOP app software. KCTU will provide Avegen with randomised stratified lists ahead of the trial start date, to which PINs will be assigned consecutively at the time of randomisation.

Randomisation will be at the individual level in the ratio 1:1:1 to the three trial arms and performed independent of the study team and automatically within the STOP app. Block randomisation with randomly varying block sizes will be used.

Randomisation stratifiers (measured at screening) will be:

1. Study site (Bath, London)
2. Gender (Male, Female)

Concealment mechanism {16b}

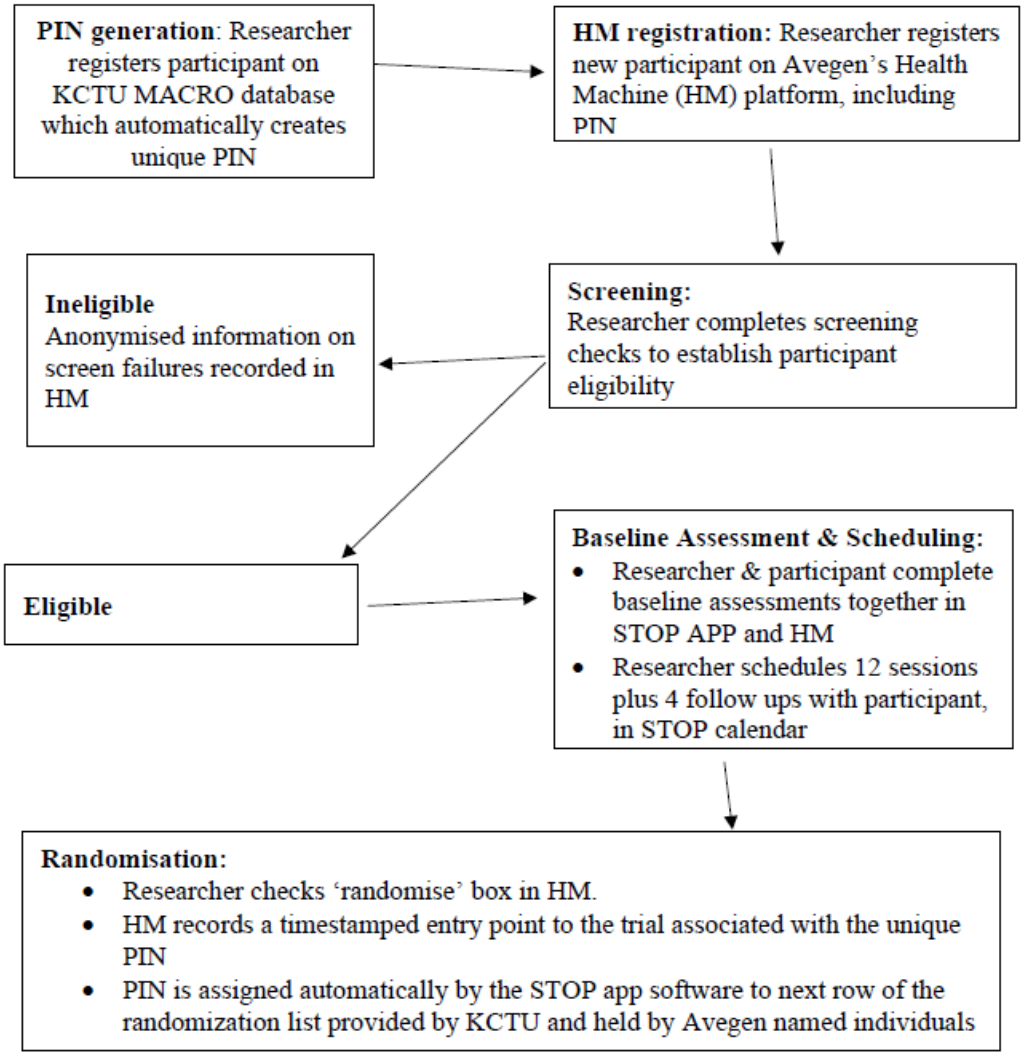
In this study randomisation lists including participant PIN numbers will be generated and communicated by KCTU to Avegen. Avegen will be responsible for implementing the randomisation process within their software platform. The patient's assignment to study arm (1, 2 or 3) will remain concealed from all those involved in the study except two nominated individuals at Avegen. These individuals will receive the KCTU randomisation lists and programme the HM/STOP app to consecutively assign PIN numbers to the randomisation list at the time the researcher completes the 'randomise' checkbox in HM. These individuals will not be part of the Avegen team working on STOP and will have no direct interactions with that team. Data output for DMCs and TSC will be generated with labels A, B and C and only the partially unblinded trial statistician will have access to them. The senior statistician will only see collapsed data.

All communications with patients about their assignment to condition (e.g. publicity, information sheet, in-app information etc.) will use the following wording: "You will be randomised to one of three different procedures, of varying lengths, some of which we think may help reduce symptoms of paranoia."

Neither patients nor researchers will be able to discern the difference between study arms since all will be procedurally identical and all will involve 12 scheduled sessions (the 6 arm intervention group will receive 12 in-app interim assessments, but without the preceding in-app therapy session)

Implementation {16c}

Figure 6. Procedure for randomisation



PIN allocation will be undertaken by trial researchers going to www.ctu.co.uk and clicking the link to access the MACRO system. A full audit trail of data entry will be automatically date and time stamped, alongside information about the user making the entry within the system.
Randomisation errors will be communicated by Avegen to the statisticians via notes added by Avegen in their system. The CI or delegate (e.g. Trial Manager) may request Avegen to add notes against individual subject entries to clarify data entry errors related to randomisation.

Blinding (Masking):
Who will be blinded {17a}

Table 3. Blinding arrangements

Fully blinded (only PIN is known OR only Arm 1,2,3 is known)	**Partially blinded** (correspondence between Arm 1,2,3 and PIN known)	**Unblinded (correspondence between 6 sessions, 12 sessions, control and PIN is known)**
All coapplicants (including PI, and lead statistician) All staff directly employed on the grant (e.g. project manager, researchers, clinicians) Other (non-designated) members of Avegen Members of McPin charity LEAP group members TSC group members All participants allocated a PIN	George Vamvakas (staff statistician)	2 designated members of the Avegen team (not permitted to have direct contact with others) DMC members, as necessary

Only two individuals will be fully unblinded throughout the trial. Two members of the Avegen team will be designated for this role and these individuals will not be permitted to have direct contact with anyone listed in the 'fully blinded' or 'partially blinded' column of the table above. These individuals are not involved in the STOP project directly and thus will have no reason to interact with the Avegen team about the STOP trial and blinding of participants.

Maintaining blinding of the study team and participants is paramount, and care will be taken to avoid accidental unblinding. Any cases of inadvertent unblinding (e.g. researcher becoming aware of the content of sessions during routine phone call with participant) will be recorded by researchers and reported to the Project Management Group (PMG). The DMC and TSC will monitor cases of reported unblinding at each centre and advise as appropriate.

If a break in blinding occurs that participant continues with their participation and sensitivity analyses will be performed to assess the impact. Blinding breaks will be recorded on a dedicated 'code break' log.

Procedure for unblinding if needed {17b}
It is not anticipated that there would be a need for emergency code break or any code break during the trial. However, if there is a compelling reason to break the blind on an individual basis (for example there is an incident of serious self-harm or suicide occurring in a study

participant, or we become aware that a study participant has committed a harmful act towards another person) then this would be done via an approach from the Chief Investigator (or delegate) to one of the two named Avegen randomisation list holders.
Where a code break is considered the first line procedure will be to:

1. Alert and seek advice from the DMC chair under chair's action.
2. Provide the site lead clinician (i.e. clinical researcher at London; site PI at Bath) with the content of the therapeutic session(s), while explaining that the patient may or may not be in that study arm. This permits the lead clinician to discuss possible triggers related to the therapy content with the patient (who will know whether they did cover those topics in the app) but allows the research team to remain blind.
3. Where the clinician or DMC chair subsequently requests a code break, this will be provided.
4. Details of any such occurrences will be reported to the Trial Steering Committee and Data Monitoring Committee at their scheduled meetings

METHODS – Data collection, management, and analysis

Data collection methods

Plans for assessment and collection of outcomes {18a}
Study researchers at both sites will be provided with appropriate training on the protocol for collecting trial data. Copies of data collection forms (i.e., case report forms, or CRFs) will be kept in the Trial Master File.

See Trial Design section {12} for an overview of the assessment schedule.

Baseline Assessment will include the following measures:

- Socio-demographic information: age, gender, ethnicity, IQ (Wechsler Test of Adult Reading, WTAR), educational level, employment and relational status, and self-reported dyslexia.
- Credibility/expectancy questionnaire items (CEQ)[49]. Three items from this measure will be used to assess participants' expectation of benefit from participation.
- Suggestibility measured using the 14-item Persuadability subscale (PER) of the Multidimensional Iowa Suggestibility Scale (MISS; [77]).
- Paranoid beliefs (please refer to the *Outcomes* section for a detailed description of each assessment)
    - Similarity rating test (SRT) [26, 27]. There are 9 SRT item sets in total, with SRT set 1 used in the screening stage; SRT set 2 and 9 counterbalanced across participants within each study arm for baseline and the four main follow-up assessments. The remaining six SRT item sets are counterbalanced across interim timepoints for each participant, within each trial arm according to a pre-defined schedule.
    - Scrambled sentences task (SST) [33, 34]. There are 15 SST item sets in total, with SST set 1 and 2 counterbalanced across participants within each study arm for baseline and the four main follow-up assessments. The remaining 13 SST item sets are counterbalanced across interim timepoints for each participant, within each trial arm according to a pre-defined schedule.

- Clinical symptoms assessed using (please refer to the *Outcomes* section for a detailed description of each assessment):
    - PANSS Item 6 (P6 Suspiciousness/Persecution)
    - Paranoia Scale
    - Green Paranoid Thought Scale Revised (R-GPTS)
    - Paranoia Worries Questionnaire
    - Hospital Anxiety and Depression Scale (HADS)
    - Structured Clinical Interview for DSM-5 (SCID-5) [50], will be conducted by a trained Clinical Psychologist during baseline assessment.
- Recovery measures (please refer to the *Outcomes* section for a detailed description of each assessment):
    - Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)
    - Questionnaire about the Process of Recovery (QPR)
    - Recovering Quality of Life (ReQoL) questionnaire.

Follow up assessments are given at 6-, 12-, 18-, and 24-weeks post-randomisation to measure longer term effectiveness of the intervention. The primary outcome is the severity of paranoid symptoms as measured by the Paranoia Scale at 24 weeks. A design overview is shown in Figure 1.

Interim assessments. The design includes every arm receiving weekly post-session 'interim' assessments for 12 weeks. This permits a dose-response analysis by allowing retrospective analysis of change in target mechanism and symptoms over time as the therapy progresses. Sessional assessments comprise a small subset of the main outcome measures as follows:

- Paranoid beliefs measured using an 8-item version of the *SRT* and the *SST* [26]
- Clinical symptoms assessed by *Paranoia Scale* [24] and *R-GPTS* [36]
- State mood change measured pre-post each session using visual analogue scales (*VAS*; anxious, sad, paranoid, friendly) [47]

Follow-up Assessment will include the full battery of outcome measures at 6-, 12-, 18-, and 24-weeks as shown in Figure 4. The Ben-Zeev et al. [46] mobile app user satisfaction measure will be given at the 24-week follow-up only. The NEQ [80] which quantitatively assesses adverse effects will be given at 12-week follow-up only.

Plans to promote participant retention and complete follow-up {18b}
As above, in 'strategies to improve adherence' section. Furthermore, remote testing is likely to increase drop-out rates but removes the risk of results being contaminated by shifting assessment mode mid protocol due to unpredictable local or national lockdowns. On balance, campus protocol is preferred due to importance of low drop out.

A bespoke schedule of email and mobile app reminders to both researchers and participants is programmed as part of the STOP app development, as detailed in the Table.

Table 4. Schedule of reminders and notifications

Reminder/Notification (sent 9am on next calendar day)	To Participant		To Researcher
	Push notification (in app)	Email	Email & HM notification
Day before a scheduled session	x	x	
Incomplete session: • session not started on scheduled day • session only partially completed on scheduled day	x	x	x
Missed session: participant omits or only partially completes scheduled session.	x	x	x
Mood worsening across 3 consecutive sessions on any VAS scale			x
Day before a scheduled follow-up (6, 12, 18, 24 wks)	x	x	x

Participants will receive £20 for each follow-up assessment they complete (4 in total) as reimbursement for their time. They will receive an additional £20 final payment upon completion of the last assessment. If travel proves necessary, we will provide a suitable allowance to contribution towards expenses for participants to physically travel to study sites.

Data management {19}
A separate Data Management Plan (DMP) for the study is available and is appended to this protocol (see Appendix 2).

The majority of study data will be collected on the STOP app or by direct entry into the Health Machine platform by researchers. The study team will access these data by regular data extracts of raw, non-aggregated data points provided by Avegen. Full data extracts will be sent automatically to a nominated study team member (George Vamvakas) by Avegen every 6 mths to prepare a comprehensive report for the DMC in order to meet DMC reporting requirements or whenever requested by DMC. In addition, data related to study progress (e.g. eligibility, recruitment, retention, drop-out etc.) and study milestones (e.g. adherence) will be provided more regularly to an agreed schedule. Data cleaning will be performed by the study team (George Vamvakas with support from other team members) after receipt of the raw data files from Avegen.

Data not captured by Avegen systems will be managed using KCTU procedures. A bespoke web based electronic data capture (EDC) system will be designed, using the InferMed Macro 4 system. The EDC will be created in collaboration with the trial analyst/s and the CI and maintained by the KCTU for the duration of the project. It will be hosted on a dedicated server within KCL.

The CI or delegate will request usernames and passwords from the KCTU. Database access will be strictly restricted through user-specific passwords to the authorised research team members. It is a legal requirement that passwords to the EDC are not shared, and that only those authorised to access the system are allowed to do so. If new staff members join the study, a user-specific username and password must be requested via the CI or delegate (e.g Trial Manager) from the KCTU team and a request for access to be revoked must be

requested when staff members leave the project. Study site staff experiencing issues with system access or functionality should contact the CI or delegate (e.g. Trial Manager) in the first instance.

Participant initials and date of birth will be entered on the EDC, NHS number, email address, participant names and addresses, and full postcodes will not be entered into the EDC. No data will be entered onto the EDC system unless a participant has signed a consent form to participate in the trial. Source data will be entered by recruiting site staff, typically within 14 days of data collection by authorised staff onto the EDC by going to www.ctu.co.uk and clicking the link to access the MACRO 4 EDC system. A full audit trail of data entry and any subsequent changes to entered data will be automatically date and time stamped, alongside information about the user making the entry/changes within the system.

The CI team will undertake appropriate reviews of the entered data, in consultation with the project analyst, for the purpose of data cleaning and will request amendments as required. No data will be amended independently of the study site responsible for entering the data.

At the end of the trial, the site PI will review all the data for each participant and provide electronic sign-off to verify that all the data are complete and correct. At this point, all data can be formally locked for analysis.

Upon request, KCTU will provide a copy of the final exported dataset to the CI in .csv format and the CI will onward distribute as appropriate.

KCTU will provide an additional specific DMP with respect to data held in the MACRO database (i.e. any data collected outside of the STOP app and Health Machine platform).

Confidentiality {27}
All information provided by participants will be kept completely confidential, within the limits of the law (namely the Data Protection Act). The limits of confidentiality and possibility of action arising from certain disclosures (e.g., participant discloses information that puts themself or someone else at risk of serious harm) will be outlined in the participant information sheets and will be discussed as part of the informed consent process.

Data confidentiality
Public sharing of data
The quantitative data generated by this study will be suitable for sharing after having been fully anonymised and will be deposited with UK Data Archive. Preparation for this purpose will be informed by Keerie et al [72], Hrynaszkiewicz et al [73] and El Emam et al [74] to reduce the possibility of re-identification. We will include a specific item in our informed consent procedures allowing participants to indicate whether they consent to anonymous data sharing via a national repository. The data deposit will be limited to those cases giving consent.

Active trial data collection
Throughout the active trial period all data collected will be kept confidential and held in pseudonymised form on secure password protected University servers. They will include

indirect identifiers (e.g., age, gender, ethnicity and sensitive clinical diagnostic information). Participants will be assigned a unique ID within the datasets. Direct identifiers (e.g. name, contact details etc.) will be stored separately by researchers in a different location and used only for the purpose of liaising with participants during their participation in the trial.
Only fully anonymised aggregated quantitative data will be used when the results of the study are published or discussed so participant identity will never be revealed in any study reports.

All qualitative data will be pseudonymised and each participant will be identified using pseudonyms to maintain confidentiality. The interviews will be audio-recorded and transcribed. Once transcriptions are available, the recordings will be destroyed. No identifiable information will appear on transcriptions (e.g. full names, exact ages, addresses, occupations). Details will be shared with participants in a separate information sheet, prior to signing a separate consent form. All participants will be made aware that pseudonymous data will be securely stored for 7 years unless they request to have the data destroyed as soon as it has been analysed. There are no plans for public deposition of qualitative data.

Plans for collection, laboratory evaluation and storage of biological specimens for genetic or molecular analysis in this trial/future use {33}
Not applicable because no biological specimens will be collected in this study.

Statistical methods

Trial objective
We are interested in the estimand for the effect of treatment under a treatment policy initially assigned at baseline as expected under routine clinical care - as defined by International Council for Harmonisation [81] - at prespecified follow-up times. Following the ITT (Intention to treat) principle (which corresponds to the treatment policy strategy approach), this requires estimating the ITT estimand, which is defined as the average treatment effect (reduction of Paranoid symptoms measured using the Paranoia Scale) between 6 sessions and 12 sessions versus control at 24-week follow-up, regardless of adherence to the allocated study treatment (or other intercurrent events that could occur) for all randomized individuals. This has the consequence that subjects allocated to a treatment arm need to be followed up, assessed and analysed as members of that arm irrespective of their compliance to the planned course of treatment [81]. This approach requires the collection of outcome data after treatment discontinuation and, we are required to treat missing data after treatment discontinuation as though they had been observed [81].

Intercurrent Events
Intercurrent events are events that occur after randomization and which either preclude the observation of the outcome or affect its measurement and/or interpretation[81,82].

Potential intercurrent events related to the disease or the specific intervention are:

- Non-adherence to treatment (treatment discontinuation, missed sessions or not starting treatment): expected to be independent of treatment arm due to blinding
- Use of non-trial treatments: Expected to be independent of treatment arm due to blinding and not problematic for this trial
- Rescue treatment and/or discontinuation due to an SAE
- Treatment switching: Not possible due to the design of the app

- Death: unlikely to be related to the trial

Handling of intercurrent events
Intercurrent events related or unrelated to the treatment or disease (e.g. moving away) still allow for collecting follow-up data for clinical outcomes, as primary and most secondary outcomes are collected through the app. We will collect follow-up data and include all data from all cases regardless of whether intercurrent events happened or not. Death is considered an unlikely event in this study.

Baseline comparability of randomised groups
A detailed description of variables measured at baseline is given under 'Methods – Baseline assessment'. Baseline variables will be described overall and by intervention group. Frequencies and proportions will describe categorical variables, while numerical variables will be described using mean and standard deviation (SD) if normally distributed and median and interquartile range (IQR) if not normally distributed.

Baseline variables to be described will be: age, gender, ethnicity, IQ (WTAR), educational level, employment and relational status, and self-reported dyslexia, the Paranoia Scale (PS), the Credibility/Expectancy questionnaire (CEQ), Paranoid beliefs measured by SRT and SST, the PANSS Item 6 (P6 Suspiciousness/Persecution), the Green Paranoid Thought Scale Revised (R-GPTS), the Paranoia Worries Questionnaire, the HADS, the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), the Questionnaire about the Process of Recovery (QPR), and the Recovering Quality of Life (ReQoL) questionnaire.

No statistical testing of the baseline differences between randomised groups will be done.

Descriptive statistics for outcome measures
The primary and secondary outcomes will be summarised at baseline, six weeks, twelve weeks and 24 weeks post-randomisation by intervention group and overall. Mean and SD or medians and IQR will be used to summarise normally distributed and non-normally distributed variables, respectively.

Inferential analysis
The analyses of effectiveness outlined in this strategy will be pragmatic, based on intention-to-treat (treatment policy strategy) and will utilise all available follow-up data from all randomised participants. All analysis will follow the intention to treat principle [51]. Group difference in mean estimates and associated 95% confidence intervals between 1) the 6-session treatment arm and control, and 2) between the 12-session treatment arm and control will be reported. These will be estimated using models that account for the repeated measurements of the participants, as described below. The main statistical analyses aim to estimate the group mean differences at the 6^{th} week (end of the 6-session treatment), the 18^{th} week, and the 24^{th} week post-randomization (6 weeks and 12 weeks post-treatment observation time).

In this study, patients are recruited from two sites (London and Bath). Randomization is stratified by site and gender, which will always be included as covariates in all statistical models. Post-randomisation measurements are taken at 6, 12, 18 and 24 weeks; we expect an increasing drop-out rate as treatment progresses and as a result, we will analyse all four time-

points simultaneously, under (restricted) maximum likelihood, to reduce potential biases and to maximise power.

Analysis of primary outcome
The primary outcome is mean paranoid symptoms measured using the Paranoia Scale (PS) [24] at 24 weeks post-randomization. PS ranges from 20 to 100 and will be treated as continuous. Our feasibility study suggested that the primary outcome is normally distributed and ranges between 20 and 80 points.

To evaluate the effect of a 6- and a 12-session STOP mobile App compared with a 12-session text reading control, a linear *mixed-model repeated measures* (mmrm) approach [52] will be fitted using restricted maximum likelihood (reml) . The model will allow a separate mean paranoia scale (PS) parameter at each assessment time in each treatment group and an unstructured covariance matrix of the response-level residuals.

Under the repeated measures setup, we will treat baseline outcome and dummy variables for the randomisation factors (study site, gender) as independent variables with a design matrix that allow no treatment difference at baseline (though a different residual variance at follow-up with non-zero covariances over time and across treatment groups). Therefore, *treatment randomisation group* (with levels denoting control, 6-sessions, and 12-sessions treatment), *time* (with levels denoting weeks 6, 12, 18 and 24 weeks), *baseline PS score, treatment by time interaction, baseline PS score by time interaction*, *study site*, and *gender* will constitute the fixed part of the model. Continuous baseline covariates will be centred at their means. This setup allows us to estimate the treatment effect at 24 weeks post-randomisation and lets follow-up data be treated as ignorable under a missing-at-random assumption. At the same time, the model accounts for the possible imbalance due to random sampling in baseline measurement of the outcome variable to control for pre-treatment differences.

The mmrm model will be fitted using Stata's mixed command as:

```
mixed PS i.time i.treatment c.baseline_PS i.time#i.treatment
i.time#c.baseline_PS i.study_site i.gender || participant_id:, nocons
residuals(unstructured, t(time) by(treatment)) reml
```

where:

PS: continuous dependent variable denoting mean paranoid symptoms, measured at weeks 6, 12, 18 and 24.

i.time: categorical independent variable indicating assessment time, coded as: 0 (6 weeks), 1 (12 weeks), 2 (18 weeks), 3 (24 weeks).

i.*treatment*: categorical independent variable denoting the treatment arm, coded as: 0 (control), 1 (6-week treatment), 2 (12-week treatment).

c.*baseline*_PS: continuous independent variable denoting mean paranoid symptoms at baseline.

i.*time*#i.*treatment*: interaction between categorical time and categorical treatment.

i.*time#c.baseline_PS*: interaction between categorical time and continuous baseline paranoia symptoms.

i.study_site: categorical independent variable indicating study site, coded as: 0 (London), 1 (Bath).

i.gender: categorical independent variable indicating gender, coded as: 0 (females), 1 (males).

participant_id: the participant's ID.

|| *participant_id*:, *nocons*: instructs Stata that the data are clustered by participant. The model allows for response-level variability only, and assumes that participant-level variability has been explained away by randomisation; the option nocons stops Stata fitting a random intercept for each participant.

residuals(unstructured, t(time) by(treatment)): estimates a residual covariance matrix based on the data at hand (unstructured) for each treatment arm separately (by(treatment)): an unstructured matrix does not impose any particular forms of correlation between the unexplained part of the responses taken at two different time points, and estimates different residual variances for each time point.

reml: instructs Stata to fit the model using Restricted Maximum Likelihood [53].

Treatment differences with 95% confidence intervals at follow-up will be calculated via Stata's lincom command, which enables estimation of linear combinations of model parameters as well as their degree of uncertainty. In addition, standardised effect sizes (Cohen's d calculated as estimated treatment difference divided by the standard deviation of baseline outcome) with 95% confidence intervals will be calculated.

Model assumption checks
The linear mixed effects models assume Normally distributed residuals. This will be checked when describing the data. Outcome residuals for each time-point will be plotted using a Q-Q plot to check for Normality and the existence of outliers. If violations of the assumptions are observed, bootstrapped 95% confidence intervals will be calculated using the bootstrap procedure for multilevel data.

Analysis of secondary outcomes
For the secondary outcomes, we will use the same model to evaluate the effect of the STOP intervention, provided that the outcome values can be assumed to be normally distributed. Where this is not the case, we will make use of generalised random effects models with robust standard errors, under the missing-at-random assumption about the data. For binary outcomes we will fit random effects logistic models with Stata's melogit command and for ordered outcomes random effects ordinal logistic models with Stata's meologit command.

Method for handling multiple comparisons
We control the type 1 error for our two planned and orthogonal contrasts (6 months against control and 12 months against control at 6 weeks follow-up) by using LSD test procedure,

which does not require a correction of the familywise error of pairwise comparisons of three groups if the main effect is significant. No correction will be made for multiple comparisons involving the primary outcome.

Besides our main outcome (Paranoia Scale), we have identified 5 key secondary outcomes (two measures of the target mechanism: SRT, SST; three measures of clinical symptoms: GPTS, HADS, PWQ) upon which we will perform hypothesis testing.

For the remaining variables, we will provide statistics and estimated treatment effects as purely explorative results and do not perform formal statistical tests. Standard errors will be provided for the reader to make their own formal assessment if desired (Senn, 2017).

Analysis of Safety

Given that there is a small likelihood that the STOP device can trigger paranoia symptoms in users over and above their baseline incidence levels of paranoia, we were classifying STOP device as a "Class B - Can cause non-serious injury". We will therefore analyse AE/SAE which involves paranoia symptoms separately from other AE/SAE to identify potential increases.

We will assess whether each AE and SAE is related to the app itself and present any AEs and SAEs classified as potentially related to the app (including all paranoia symptoms) in the following ways:

- Each event, date and reason for potential relation to the app will be listed individually
- We will then i) plot the number of adverse events and SAEs per predefined time period (e.g. 6 x 4-week blocks) and ii) produce a cumulative plot over time. The same descriptive analyses will be presented per arm. The events will be presented as collapsed AE and SAE as well as grouped into main categories. The individual events and numbers will be regularly assessed by the PI and by independent members of the DMC at each meeting. The DMC will also assess any critical development between bi-annual meetings when necessary.

We do not expect the number of potential app-related AEs/SAEs to occur frequently enough to perform formal statistical analyses during the trial, but for the final report we will report the incidence rate (total number of those having the event divided by the person-months at risk) and the ratio of incidence rates between the two treatment arms per time period will be calculated to allow detection of any safety concerns within the treatment arm.

Qualitative analysis. The data from the adverse events checklist in the 12 session arm will be analyzed using qualitative content analysis to produce a list of adverse events along with their frequencies, seriousness, relatedness to the app, and possible methods of prevention/mitigation. Additionally, we will examine the demographic and clinical characteristics of those who experienced adverse events to identify patients who might be at a higher risk.

The NEQ data are provided as a total/overall score of the negative effects that are related to treatment and those that are related to other circumstances. As these data provide quantitative

scores for every participant we will be able to perform comparative statistical tests between treatment and control arms using a linear regression approach.

Missing data

Minimizing attrition at follow-up

We will try to minimize attrition at follow-up by reminding participants the day before their scheduled follow-up measurements by email and by using automatically generated reminders on their mobile phones. Also, researchers will be present at follow-up appointments and will be aware of the need to keep data loss to a minimum.

Reporting of missingness

We will follow CONSORT guidelines of reporting the reasons why patients were lost of follow-up and hence why outcomes are missing.
We do not expect a differential loss of participant counts between arms because participants will not be aware if they are randomized into a treatment or control group. Therefore, bias due to missingness should be similarly distributed among treatment arms. This assumption will be assessed by comparing the number of participants between arms.
To identify the potential for bias due to missing data, we will present a table comparing the distribution of baseline data by treatment arm for patients with observed follow-up data and for patients with missing follow-up data separately for each study arm (Schultz et al 2010).

Missing data will be described both as a proportion in each variable separately as well as using a graph that ranks variables from most complete to least complete. This will be done for all participants together as well as for each arm separately. Due to the data collection via app, we expect that most clinical outcome data are either all collected for a participant or all are missing. Given the amount of missing data we encountered in the feasibility study, we anticipate a high degree of completeness.

Handling missing data in the analysis

If there is complete follow-up, an estimand based on treatment policy can be estimated with little assumptions. However, despite efforts to collect data on all randomized participants, invariably there will be some missing data. We do not expect missing data at baseline. Missingness at the baseline of the main clinical outcome variables is not possible due to the recording method in the app.
However, some subjects may discontinue from the trial and others may miss one or two of the three follow-up visits. Reasons for these events will be recorded as detailed as possible. The missing data will be addressed in the statistical analyses using principled methods such as Maximum Likelihood, which can provide an unbiased and precise estimate of the ITT estimand in the presence of missing data at random.

When missing data exist, analyses rely on assumptions about the behaviour of the participants after dropping out. A robust estimate can only be made if the assumptions need to be made are justifiable and plausible and hence assumptions need to be explicitly stated. However, the assumptions cannot be proven and as a result, sensitivity analyses will be performed to explore the robustness of the inference to these assumptions.

Missing data in baseline variables
If the proportion of missing data in baseline variables is small (approximately 1-2%) we will fill the missing data back in with the mean of the baseline variable [54]. If the proportion of missing data in the baseline variables is large, we will consider using multiple imputation [55] by chained equations, which can accommodate the imputation of missing data in baseline variables as well as in outcome measures at the same time. Results from this process will be contrasted with results from the complete case data analyses.

Missing data in the outcome

The described mixed model will be estimated via restricted maximum likelihood that is valid under the missing at random (MAR) assumption and allows to include all patients with at least one follow-up observation. The MAR assumption requires that all predictors of the missingness mechanism be included in the models to maximise the likelihood that, conditional on these predictors, the outcome is missing at random.

To make the assumption of missing at random more plausible, we have prespecified baseline variables that are assumed to be a potential predictor both of the incomplete variables and of the probability of a value being observed. These prespecified variables are age, gender, ethnicity, IQ (WTAR), educational level, employment and relational status, self-reported dyslexia, Paranoid beliefs, Paranoia Scale, and HADS. For the primary analysis we will exclude Paranoia Scale from this list, since it will be used as a dependent variable. We will perform logistic regressions (with missing at follow-up (yes/no) as dependent variable) to identify key variables among the prespecified variables that are predictive of missingness in the analyses following the guidelines of Carpenter and Smuk [83].

To assess whether missing outcome data are predicted by baseline variables, we will firstly construct three binary (0/1) indicator variables per outcome, capturing whether any of the outcomes are missing their 6-, 18- or 24-week value, respectively. These indicators will be the dependent variable in logistic regression models that include intervention group, PS score at baseline, gender and study site as independent variables. We call these models "core missingness" models. Then, we will examine whether several baseline variables (see next paragraph) correlated with the outcome of interest, can predict missingness in the outcome when added to the core model, by following two steps. In the first step, we will add the baseline variables in the core model, one at a time. They will be considered to predict missingness if there is a significant relationship at a 10% level. In the second step, the significant variables identified in step one, will enter a new model all at the same time. The new model will be a logistic model with the same binary indicator of "missingness" used in step one. The statistically significant variables from this will constitute the predictors of missingness and will be incorporated in the models for the analysis of the primary or secondary outcomes as part of a sensitivity analysis. Interactions between covariates and time will be assessed relaxing the assumptions that missingness patterns are similar between the treatment arm and over time. We will report how sensitive our results are to the inclusion of these variables and post-treatment group difference estimates and associated confidence intervals will be reported.

Sensitivity of results due to missing data by assuming missing not at random (Primary Analysis)

Analysis of data where the outcome is incomplete always requires untestable assumptions about the missing data commonly that the data are missing at random. The estimand would be inappropriate if a large proportion of participants leave the study due to intercurrent events and are not available for the final endpoint measurement. In this case, we assume that the data are potentially not missing at random (MNAR). To address this the primary analyses will be conducted using a pattern mixture with multiple imputations model [86]. This approach formulates the MNAR problem in terms of the different distributions between the missing and observed data and different distributions can be specified for different patterns of missingness of (groups of) participants that reflects a specific assumption appropriate to their treatment arm, drop-out time and possibly other relevant information.

We will perform sensitivity analyses to explore the effect of departures (varied over a plausible range) from the assumption of missing at random made in the main analysis following the steps recommended by White et al [84].
We will set up a multiple imputation model consistent with the primary substantive analysis model. We will use multiple imputations to impute missing values under a MAR assumption and modify the MAR imputed data to reflect a plausible range of patterns related to the effect of non-compliance followed by drop-out following the guidelines of Carpenter et al [86].

We will explore three types of information-anchored sensitivity analysis:

1. One approach acknowledges that it is possible that drop-out is caused by the app itself and not the treatment (we assume that participants do not know if they are in treatment or control conditions) and the app itself may have a negative influence. To examine this a delta-adjusted approach will be performed, where the MAR imputed values are modified to reflect the assumed MNAR mechanism [87]. Participants who discontinue the treatment become worse by some fixed amount compared with participants who continue the trial. After imputing the data under MAR, we will increase the imputed values by 2%, 5%, 7%,… of the treatment group means at the main time point, until we reach the 'tipping point' where the treatment effect is no longer significant. A-priori, we believe that a 5% difference is plausible so that if the test for treatment effect remains significant, we will report the results are 'robust to plausible a-priori agreed departure from MAR.'
2. Another approach is the copy reference-based pattern mixture model which assumes that the post-dropout response profile of participants who discontinue the randomized treatment will be like that of patients on the TAU arm. We will use 'Copy Reference' as a more extreme sensitivity analysis [86], where as reference we will choose the control arm. Under Copy Reference when participants drop out of the study the means and the covariance matrix of their response distribution, both before and after drop-out, are replaced entirely with those from the reference arm. Hence, if the reference group is assumed to be the control arm, then this assumption mimics the case where those dropping out do not respond to treatment. For participants in the reference arm, their imputation model is that of the MAR assumption.
3. A third approach is to assume that the app and the treatments have additive negative influences on the participants which result in drop-out. Here we would assume differential worsening of the clinical outcome by the different arms.

Terminal Events
Death is classified as an intercurrent event but the treatment policy strategy cannot be implemented for intercurrent events that are terminal events, since values for the variable after the intercurrent event do not exist. In the unlikely event of a death, we will, therefore, expand our sensitivity analysis to include a scenario where the terminal event would not occur (i.e., the life of the participant was saved) and add the value which the clinical outcome would have taken in the hypothetical scenario defined (i.e., getting worse). These scenarios can be added to any of the three pattern mixture models described above, as additional instances (patterns) that assume extreme worsening of outcome.
Finally, if death is related to the treatment and/or the app, our treatment policy strategy estimand will not be relevant anymore as this would be reported as a serious adverse event that would prompt external investigations into the continuation of the trial.
Any deaths confirmed as unrelated to the treatment and/or the app will be assumed to be MAR cases, but MNAR sensitivity analyses will still be performed.

All scenarios will be prespecified in the analyses and the range of shifts of imputed values will be guided by content experts. Additional scenarios will be performed on the request of the DMC committee. We will then analyse the data of the different models as per our primary analyses following the standard recommendation of multiple imputation guidelines to obtain a single MNAR estimate and standard errors and report any changes in the estimated treatment effects.

Missing items in scales and subscales
This paragraph is not relevant for the primary outcome variable but only for some of the secondary outcome and baseline variables which are based on questionnaires or similar item-based measurement scales.

The planned strategy for handling missing data at the item and scales will depend on the amount of missing data observed and the planned analyses for the outcomes.

Missing item data will be imputed using pro-rating that is, by replacing the missing item score by the mean of the observed items if less than 20% of the items scores are missing. Items within each scale are indicators of a specific concept and as a result assumed to be closely and positively correlated and therefore regarded as a particularly applicable technique [56, 57]. Simulation studies have shown that pro-rating (or case mean substitution) is a robust method when data are missing on less than 20% of items in both random and systematic patterns [57].

Pro-rating is implemented across items within a scale, or subscale, for each assessment and participant. In the unlikely event that more than 20% missing items on an item variable data is collected we will treat it as a missing data point.

To ensure the same strategy is followed across all scales reported in the principal paper(s), any guidance given by authors of validated questionnaires will supersede the methods outlined herein.

Comment

We did not plan a multiple imputation analysis because we intend to include all major baseline covariates which are putatively predictive of missingness and outcome. Multiple imputations would have only been superior to our proposed analysis if the imputation model could have been supplemented with additional post-randomisation variables or other baseline variables not included in the substantive analysis, which help to alleviate the biases caused by missing data. This is, however, not the case because we do not plan to collect such variables.

Sensitivity of results to potential imbalance of baseline characteristics
The randomisation of a relatively large sample of participants into groups (anticipating approximately 90 participants per treatment group in this study) ensures that the average values of individual's characteristics at baseline are similar across the groups that is, they are balanced. Imbalances, however, are likely to occur by chance and when they do, they affect the inference about the treatment effect.

In this study, the following baseline variables will be examined for chance imbalances: 1) education level, 2) age, 3) ethnicity, 4) paranoia severity, and 5) bias score (SRT). To assess the effect of potential imbalance all five variables will be controlled for and included as additional covariates in the models. Any changes in treatment differences will be reported.

Sensitivity of results to the increase in precision
The five variables mentioned in the previous paragraph were chosen because they were thought to be important in predicting the outcome and increase the precision of the treatment effects. To assess the sensitivity of the standard errors in the presence of these variables, we will re-run and report the findings after adding 1) education level, 2) age, 3) ethnicity, and 4) SRT to the primary model, and 1) education level, 2) age, 3) ethnicity, 4) SRT, and 5) paranoia severity to the secondary models.

Sensitivity analysis due to Protocol violations
A protocol violation or non-compliance is any deviation from the protocol.

The statistical analyses are primarily concerned with two protocol violations:

1) Violations with regards to the occurrence of a session and the extend of the adherence to the study, as described in Section Strategies to Improve Adherence to Interventions.
2) Violations due to unblinding.

We will assess violations to the protocol by assessing the violations (i.e. unblinding or data was collected outside the collection time frame) in further sensitivity analysis by supplementing the model of the primary analysis, with an additional binary indicator variable coded 1 for existence of any Protocol violation (i.e. unblinding at follow-up time), and 0 otherwise. The analysis of this model will give intervention effect estimates adjusted for potential effects of Protocol violation. We will report the changes in the predicted outcome differences alongside the main analysis.

Interim analyses {21b}

No interim analysis is planned.

If recruitment, adherence or other milestones are not met the DMC and TSC may request interim analyses to help inform their recommendations to the funder (MRC) about whether the study should stop or continue.

Subgroup analyses {20b}
No subgroup analyses are planned.

Explorative dose-response analyses
The study design includes 12 weekly post-session interim assessments. This permits us to develop a dose-response model to examine the effect of treatment on paranoia growth by allowing a retrospective analysis of the primary outcome over time as the therapy progresses.

The dose-response analysis is described fully in the Statistical Analysis Plan (SAP). We report here a summary of the main points. The dose-response model will be a non-linear (in the parameters) repeated-measures model. The non-linear function will be a three-parameter logistic growth function [58], based on the data we obtained during the feasibility study. This will allow us to estimate 1) the average paranoia score at increasing doses (or numbers of session), 2) the dose (or session) at which half the average paranoia score is reached and 3) the number of weeks it takes for paranoia to move from 50% to about 73% of its growth. To explore how treatment affects any of these quantities, we will incorporate into the model the treatment arm, as well as the controlling covariates (baseline outcome, study site, and gender), and random intercept parameters.

The dose-response model will be fitted using R's interface to JAGS package *rjags* [59], a program for the analysis of Bayesian hierarchical models using Makrov Chain Monte Carlo simulation.

As a sensitivity analysis, we will consider using different non-linear functions that may describe the idiosyncrasies of the sample's growth better. The choice of such functions will be informed by randomly selecting a sub-sample of study participants (approximately 10-20% of the total sample) and fitting a non-linear model to each one of these participants separately, using the R package *drc* [60]. If we find evidence of a better fit with a different non-linear function (assessed mainly through graphical means), we will change the mean of the Normal likelihood of the main Bayesian model to accommodate the new function and compare the models using the deviance information criterion.

Statistical Analysis Plan (SAP)
A detailed statistical analysis plan will be produced and submitted to the DMC and TSC for final approval.

Software
Data will be exported from the KCTU MACRO database into Stata file format (.ana). All data processing and statistical analyses for the main trial paper will be performed using Stata versions 16 or higher with the exception of the dose-response analyses which uses R.

Plans to give access to the full protocol, participant level-data and statistical code {31c}
The full protocol, including detailed Statistical analyses plan, participant information sheets, and data files will be deposited at the ISRCTN registry (https://www.biomedcentral.com/p/isrctn). This is a clinical trial registry recognised by WHO and ICMJE that accepts planned, ongoing or completed studies of any design. It provides content validation and curation and the unique identification number necessary for publication. All study records in the database are freely accessible and searchable. The statistical code will be deposited in GitHub (www.github.com), a Git repository hosting service. A trial protocol with the statistical analyses plan as an addendum will be published.

METHODS – Monitoring

Data monitoring

Composition of the coordinating centre and trial steering committee {5d}
The design, development and implementation of the trial will be managed by the Principal Investigator (PI) and the co-applicants, in consultation with service users and other expert research collaborators from within and outside of the PI's institution. A dedicated Project Manager, reporting to the PI, will oversee trial conduct, assist with governance and regulatory matters, and provide quality control.

The PMG will meet quarterly to ensure the project remains focused on delivering against its objectives and will monitor the project against its stated budget, schedule, milestones and risks. Its membership will include the investigators, the Project Manager, a junior statistician, a representative of the PI's Technology Transfer Office, and a representative of the external vendor contracted to develop the STOP app. It will oversee the preparation of reports to the Funder (MRC), TSC and DMC.

The purpose of the TSC is to oversee the trial on behalf of the Sponsor and Funder; to monitor adherence to the protocol and statistical analysis plan; to ensure the rights, safety and well-being of the participants are given the utmost consideration; and to provide independent advice on all aspects of the trial. It will consist of five members independent of the trial team (a chairperson, a statistician, two clinical academics, and a PPI representative) and three members from the senior trial team (Principal Investigator, trial statistician, and an additional PPI representative).

Patient and Public Involvement (PPI) strategy
User input is integral to, and essential for, this project. A Lived Experience Advisory Panel (LEAP) will contribute to the design and branding of STOP and will input significantly to app development. In addition, LEAP will meet quarterly to receive progress updates and provide input including to: consent form and participant/carer information sheets; website; recruitment strategies; access to user networks; publicity and dissemination and any other issues that may arise. We will specifically aim to develop LEAP members and recognise their input at every opportunity. We will offer LEAP members acknowledgement and authorship (as appropriate) in peer reviewed papers.

An independent PPI member will sit on the TSC.

Composition of the data monitoring committee, its role and reporting structure {21a}
The DMC will consist of three independent members: a chairperson, a clinical expert, and a statistician. The DMC will monitor the data emerging from the trial as they relate to data quality and the safety of participants and may request for interim analyses of trial data to be undertaken. It will make recommendations to the TSC on whether the trial should continue, stop or be modified. A copy of the DMC Charter will be available upon request from the lead investigator, J. Yiend.

Adverse event reporting and harms {22}
Full details of adverse event definitions, reporting and procedures are documented in STOP's Adverse Events SOP. Key information is listed here for easy reference.

Defining
A Serious Adverse Event (SAE) is any adverse event that:

- results in death
- is life-threatening* (i.e., at-risk of death)
- requires hospitalisation or prolongation of existing hospitalisation
- results in persistent or significant disability or incapacity
- jeopardises the patient, or requires intervention to prevent one of the outcomes listed above

*Life-threatening, by definition, refers to an event in which the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe.

Device deficiency is defined as inadequacy of the investigational medical device (the STOP app) related to its identity, quality, durability, reliability, safety or performance. This may include malfunctions, use error, or inadequacy in the information supplied by the manufacturer.

Collecting
Information concerning adverse events will be collected using an **Adverse Events Checklist.** Researchers will use this checklist to proactively enquire about patient safety at every scheduled contact (i.e. fortnightly phone call; follow up assessments; weekly phone call for those identified as high on susceptibility).

Assessing
Every AE reported by participants and captured on the checklist will be assessed for seriousness, relatedness, severity and whether it was anticipated or not, by site researchers initially, following the definitions, categorisations and procedures outlined in the SOP. 'Seriousness' is defined as above. 'Relatedness' potentially includes any event listed on the STOP Hazard Log, as well as any other event deemed by the research team to be directly related to the device or procedures of the trial.

'Anticipated' events are those captured under the 11 categories below, specific examples of which are given in the Adverse Events Checklist and the STOP Hazard Analysis.
The categories comprise an a priori definition of anticipated adverse events, have been approved by the DMC and were informed by a literature review which highlighted the

consensus statement of Rozental and colleagues [75], the adverse effects of psychological therapy reported by Mc Glanaghy and colleagues [62] and the adverse effects of psychological interventions reported by Gullickson and colleagues [76]. A further multidisciplinary workshop followed by a consensus opinion exercise was conducted. The combined result of these activities produced the following 11 categories of anticipated adverse events:

1. Technical Malfunction
2. Symptom exacerbation/ deterioration
3. Somatic/ physical effects
4. Practical burden
5. Novel symptoms
6. Lack of adherence to treatment (also includes lack of response or engagement for other reasons & withdrawal)
7. Harmful behaviours
8. Data Breach/ Privacy
9. Content/Format problems
10. Clinical care (including changes to TAU; engagement with TAU; crisis interventions)
11. Assessment Related

These categories are used in a coordinated manner to conceptualize the approach to safety across all relevant documentation including the Adverse Events Checklist for data collection and the Adverse Events Log for data recording and the STOP Hazard Analysis for risk assessment.

Recording
Details of adverse events will be captured on the Adverse Events Checklist located within the relevant Case Report Forms. Data will be transferred to the customised STOP study MACRO database provided by KCTU. The adverse events section of the database is configured from the Adverse Event Log and will record the following fields: Adverse event description (free text); start date; status (ongoing or ended), stop date (if not ongoing), severity (mild/ moderate/ severe), relatedness (not related/unlikely/ possibly/ probably/ definitely causally related to the intervention), seriousness (either not serious or coded against the criteria for a SAE)

Reporting
The study will follow the guidance given by MEDDEV 2.7/3 Rev 3 Clinical Investigations: Serious Adverse Event Reporting (https://ec.europa.eu/docsroom/documents/16477/attachments/1/translations) [61] and by the HRA (https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/safety-reporting/)

Urgent reporting. If an urgent safety measure is required in order to protect research participants against any immediate hazard to their health or safety, this will be implemented without prior authorisation from a regulatory body. The main Research Ethics Committee (REC) and the MHRA will be notified immediately and in any event within three days, in the form of a substantial amendment, that such measures have been taken and the reasons why

Any event identified as a SUSAR (suspected unexpected serious adverse reaction) will trigger the use of a product 'recall-switch' that will disable the device immediately for all users of the device (see Device Withdrawal Procedures)

All serious adverse events or device deficiencies that could have led to a serious adverse event are reportable to MHRA. Each time a new reportable event (or a new finding related to an already reported event) occurs within the trial the following actions will be taken:

- The CI will call the MHRA's medical devices adverse incident centre on 020 3080 7080 within 2 calendar days of awareness of the incident
- The MHRA SAE reporting form will be updated and emailed to MHRA's Adverse Incident Centre at aic@mhra.gsi.gov.uk

All HRA and other reporting procedures (including notification to Sponsor, trial committees and local research directorates) are outlined in the SOP. Where the event is serious and related to the trial or device it will be reported to the REC (and MHRA) immediately (and always within 15 days of the CI becoming aware).

Regular Reporting. All AEs and SAEs (from each site) will be pooled and reported to each meeting of the Data Monitoring Committee (DMC), who report to the Trial Management Committee. These aggregated statistical reports will be generated from the study's MACRO database via the Adverse Event Log or at any time at the request of the DMC Chair.Full reporting procedures are outlined in the study's Adverse Event SOP. These include an annual report to the REC and immediate report to the MHRA and chair of the DMC of every individual serious adverse event (related or not), as outlined under urgent reporting, above. The DMC will be responsible for investigating further if there are any concerns about unexpectedly high rates of SAEs, which may include being unblinded as to trial condition or seeking further data on adverse events and will advise the TSC on any ethical or safety reasons why the trial should be prematurely ended. The Funder will immediately be notified on receipt of any information that raises material concerns about safety or efficacy, and of any recommendations from the DMC to end the trial.

In the STOP study there is a potential risk of participants becoming distressed while completing the sessions, although this potential AE has been assessed as low risk in the STOP Hazard Analysis. We have data from the feasibility trial showing that participants' immediate mood was either unaffected, or slightly improved after each session. Nevertheless, the safety of the intervention will be monitored closely in three ways:

1) by measuring mood within the app (variety of mood measures pre-post each session). Follow up contact with a researcher will occur if a participant's mood worsens on 3 consecutive occasions on any state mood (VAS) scale.
2) By fortnightly (weekly for those identified as highly suggestible) semi-structured interview phone calls which will include questions about events that have occurred in the preceding week and the Adverse Events Checklist.
3) A clinical support email and study helpline will be available to participants throughout the duration of the trial.

Frequency and plans for auditing trial conduct {23}

The TSC and the DMC are responsible for ensuring the trial is conducted to a high standard. It is anticipated that these committees will convene on an annual basis, or more frequently if required (e.g., to review adverse events). The trial statistician will prepare a comprehensive report for each DMC meeting.

ETHICS AND DISSEMINATION

Plans for communicating important protocol amendments to relevant parties (e.g., trial participants, ethical committees) {25}
Any amendments to the protocol will be submitted to the REC and the TSC for approval. Subsequent changes will then be communicated to participants, as appropriate, and recorded in the ISRTCN registration.

Dissemination plans {31a}
Results will be published in good quality academic journals and presented at appropriate national and international conferences, for example the annual British Association of Behavioural and Cognitive Psychotherapies conference, or the triennial World Confederation of Behavioural and Cognitive Psychotherapies (https://wccbt.org).

The study website (www.STOPtrial.co.uk) will be used to communicate study progress and key findings.

The study and its findings will also be publicised through the professional dissemination services provided by The McPin Foundation (https://mcpin.org/), through the dissemination of materials such as infographics, leaflets, podcasts, and articles for newsletters and popular press. These will be made available on the study website and will be proactively disseminated through user networks, relevant mental health charity websites and newsletters, and NHS trust communications.

DISCUSSION

Persecutory delusions are at the severe end of a paranoia spectrum, associated with poorer health, emotional wellbeing and social functioning [2], and for which treatments need to be significantly improved. There is evidence that persecutory delusions are maintained by underlying cognitive biases, which perpetuate distressing paranoid beliefs [63], specifically biases in the interpretation of emotional ambiguity [7, 64, 65]. Consequently, there are calls for new psychological treatments to be developed which target these maintaining mechanisms [63]. To date there has been no intervention that uniquely targets the biased interpretation of emotional ambiguity that is known to be associated with clinical [64] and non-clinical [7] paranoia.
The current study will build on the previous CBM-pa feasibility study by testing a novel, entirely self-administered digital therapeutic for paranoia called STOP ('Successful Treatment of Paranoia'). STOP is the successor to CBM-pa that includes a 12-session mobile app with illustrations for each item used in the study. If effective, STOP could offer a new intervention to complement other psychological treatments (including CBT) [66]. New targeted interventions are required because CBT effect sizes for delusions are small to modest [67], a significant proportion of patients do not respond fully and resource limitations restrict availability [68]. This has been recognised by the advent of the IAPT-SMI (Increasing Access to Psychological Therapies for Severe Mental Illness) programme. The development of an evidence-base for psychological treatments is seen as a priority for IAPT- SMI [69] and the

evaluation of STOP could contribute to these developments, as a potential low-cost, accessible and flexible intervention.

TRIAL STATUS

Development and piloting of the App commenced in March 2021 and will continue until January 2022. Recruitment will commence in July 2022 and will continue until November 2023.

ABBREVIATIONS

AE: Adverse Event
SAE: Serious adverse event
CBM-pa: Cognitive Bias Modification in paranoia
CBT: Cognitive behaviour therapy
DMC: Data Monitoring Committee
LEAP: Lived Experience Advisory Panel
MRC: Medical Research Council
NHS: National Health Service
PMG: Project Management Group
PPI: Patient and Public Involvement
REC: Research Ethics Committee
RCT: Randomised controlled trial
STOP: Successful Treatment of Paranoia
TAU: Treatment as usual
TSC: Trial Steering Committee

DECLARATIONS

Acknowledgements

A Lived Experience Advisory Panel (LEAP) for the study is provided by the mental health research charity, The McPin Foundation (www.mcpin.org). The LEAP for the project is coordinated by the study co-applicant, Dr Thomas Kabir. He manages a LEAP of six people to work on the study. All members of the LEAP are invited to attend Trial Management Group meetings.

Authors' contributions {31b}

JY is the chief investigator of the study, lead the writing of the study protocol, is the lead for the trial, and will lead the interpretation and reporting of the findings.
CWH contributed to writing the study protocol.
TK is the lead for patient and public involvement in the study and contributes to all aspects of study.
JK:
PM:
EM:
EP:
TR contributed to writing the study protocol, project management and set-up, and recruitment.
SS:
DS: is the trial statistician, carried out the power calculation, assisted with study design decisions and provides statistical support for design and implementation of analyses.
GV:

PJ:
All authors read and approved the final manuscript.

Funding {4}
This research was funded by the Medical Research Council (MRC) Biomedical Catalyst: Developmental Pathway Funding Scheme (DPFS), MRC Reference: MR/V027484/1.
The Funder provided only financial support for the conduct of the research and did not have a role in collection, management, analyzation, and interpretation of data, or in writing of the manuscript.

Availability of data and materials {29}
An anonymised version of the data generated and/or analysed during this study will be available via the UK Data Archive upon written request, subject to independent review, following the publication of the main findings. A data sharing agreement would be signed by both parties, outlining the new purpose for which the data will be used, conditions of use, appropriate acknowledgement of original funding source, etc.

Ethics approval and consent to participate {24}
The study has been approved by the London-Stanmore Research Ethics Committee (reference: 21/LO/0896). We will/have obtain(ed) informed consent from all participants in the study.

Consent for publication {32}
Not applicable.

Competing interests {28}
The authors declare that they have no competing interests and are independent of the contractors.

Authors' information

Affiliations
Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
Jenny Yiend, Che-Wei Hsu, Philip McGuire, Elias Mouchlianitis, Tanya Ricci, Sukhwinder Shergill

The McPin Foundation, London, UK
Thomas Kabir

Department of Informatics, King's College London, London, UK
Jeroen Keppens

Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
Emmanuelle Peters

Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
Daniel Stahl

Clinical Trials Unit, King's College London, London, UK
George Vamvakas

Department of Psychology, University of Bath, Bath, UK
Pamela Jacobsen

Corresponding author
Correspondence to J. Yiend

REFERENCES

1. Perälä J, Suvisaari J, Saarni SI, Kuoppasalmi K, Isometsä E, Pirkola S, Partonen T, Tuulio-Henriksson A, Hintikka J, Kieseppä T. Lifetime prevalence of psychotic and bipolar I disorders in a general population. Arch Gen Psychiatry. 2007;64(1):19-28. doi:10.1001/archpsyc.64.1.19

2. Freeman D, McManus S, Brugha T, Meltzer H, Jenkins R, Bebbington P. Concomitants of paranoia in the general population. Psychol Med. 2011;41(5):923-36. doi:10.1017/S0033291710001546

3. Freeman D, Garety PA, Bebbington PE, Smith B, Rollinson R, Fowler D, Kuipers E, Ray K, Dunn G. Psychological investigation of the structure of paranoia in a non-clinical population. Br J Psychiatry. 2005;186(5):427-35. doi:10.1192/bjp.186.5.427

4. Bentall RP, Rowse G, Shryane N, Kinderman P, Howard R, Blackwood N, Moore R, Corcoran R. The cognitive and affective structure of paranoid delusions: a transdiagnostic investigation of patients with schizophrenia spectrum disorders and depression. Arch Gen Psychiatry. 2009;66(3):236-47. doi:10.1001/archgenpsychiatry.2009.1

5. Birchwood M, Trower P. The future of cognitive–behavioural therapy for psychosis: not a quasi-neuroleptic. Br J Psychiatry. 2006;188:107-8. doi:10.1192/bjp.bp.105.014985

6. Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE. A cognitive model of the positive symptoms of psychosis. Psychol Med. 2001;31(2):189-95. doi:10.1017/s0033291701003312

7. Savulich G, Freeman D, Shergill S, Yiend J. Interpretation biases in paranoia. Behav ther. 2015;46(1):110-24.

8. Freeman D, Garety PA, Kuipers E, Fowler D, Bebbington PE. A cognitive model of persecutory delusions. Br J Clin Psychol. 2002;41(Pt 4):331-47. doi:10.1348/014466502760387461

9. Wessely S, Buchanan A, Reed A, Cutting J, Everitt B, Garety P, Taylor P. Acting on delusions. I: Prevalence. Br J Psychiatry. 1993;163:69-76. doi:10.1192/bjp.163.1.69

10. Castle DJ, Phelan M, Wessely S, Murray RM. Which patients with non-affective functional psychosis are not admitted at first psychiatric contact? Br J Psychiatry. 1994;165(1):101-6. doi:10.1192/bjp.165.1.101

11. Johnson J, Horwath E, Weissman MM. The validity of major depression with psychotic features based on a community study. Arch Gen Psychiatry. 1991;48(12):1075-81. doi:10.1001/archpsyc.1991.01810360039006

12. Goodwin FK, Jamison KR. Manic-depressive illness: bipolar disorders and recurrent depression. New York: Oxford University Press; 2007.

13. Hamner MB, Frueh BC, Ulmer HG, Arana GW. Psychotic features and illness severity in combat veterans with chronic posttraumatic stress disorder. Biol Psychiatry. 1999;45(7):846-52. doi:10.1016/s0006-3223(98)00301-1

14. van Os J, Verdoux H, Maurice-Tison S, Gay B, Liraud F, Salamon R, Bourgeois M. Self-reported psychosis-like symptoms and the continuum of psychosis. Soc Psychiatry Psychiatr Epidemiol. 1999;34(9):459-63. doi:10.1007/s001270050220

15. Appelbaum PS, Robbins PC, Roth LH. Dimensional approach to delusions: comparison across types and diagnoses. Am J Psychiatry. 1999;156(12):1938-43. doi:10.1176/ajp.156.12.1938

16. van der Gaag M, Valmaggia LR, Smit F. The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: a meta-analysis. Schizophr Res. 2014;156(1):30-7. doi:10.1016/j.schres.2014.03.016

17. Mehl S, Werner D, Lincoln TM. Does Cognitive Behavior Therapy for psychosis (CBTp) show a sustainable effect on delusions? A meta-analysis. Front Psychol. 2015;6:1450. doi:10.3389/fpsyg.2015.01450

18. Turner DT, van der Gaag M, Karyotaki E, Cuijpers P. Psychological interventions for psychosis: a meta-analysis of comparative outcome studies. Am J Psychiatry. 2014;171(5):523-38. doi:10.1176/appi.ajp.2013.13081159

19. Elkis H. Treatment-resistant schizophrenia. Psychiatr Clin North Am. 2007;30(3):511-33. doi:10.1016/j.psc.2007.04.001

20. Moritz S, Woodward TS. Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. Curr Opin Psychiatry. 2007;20(6):619-25. doi:10.1097/YCO.0b013e3282f0b8ed

21. Waller H, Freeman D, Jolley S, Dunn G, Garety P. Targeting reasoning biases in delusions: a pilot study of the Maudsley Review Training Programme for individuals with persistent, high conviction delusions. J Behav Ther Exp Psychiatry. 2011;42(3):414-21. doi:10.1016/j.jbtep.2011.03.001

22. Yiend J, Trotta A, Meek C, Dzafic I, Baldus N, Crane B, Kabir T, Stahl D, Heslin M, Shergill S, McGuire P, Peters E. Cognitive bias modification for paranoia (CBM-pa): Study protocol for a randomised controlled trial. Trials. 2017;18(1):298. doi:10.1186/s13063-017-2037-x

23. Leung CJ, Fosuaah A, Frerichs J, Heslin M, Kabir T, Lee TMC, McGuire P, Meek C, Mouchlianitis E, Nath AS, Peters E, Shergill S, Stahl D, Trotta A, Yiend J. A qualitative study of the acceptability of cognitive bias modification for paranoia (CBM-pa) in patients with psychosis. BMC Psychiatry. 2019;19(1):225. doi:10.1186/s12888-019-2215-3

24. Fenigstein A, Vanable PA. Paranoia and self-consciousness. J Pers Soc Psychol. 1992;62(1):129-38. doi:10.1037//0022-3514.62.1.129

25. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13(2):261-76. doi:10.1093/schbul/13.2.261

26. Mathews A, Mackintosh B. Induced emotional interpretation bias and anxiety. J Abnorm Psychol. 2000;109(4):602-15.

27. Eysenck MW, Mogg K, May J, Richards A, Mathews A. Bias in interpretation of ambiguous sentences related to threat in anxiety. J Abnorm Psychol. 1991;100(2):144-50. doi:10.1037//0021-843x.100.2.144

28. Department for Constitutional Affairs. Mental Capacity Act 2005, Code of Practice. United Kingdom: Crown Copyright; 2007.

29. Jacobsen P, Wood L. Risk of contamination when planning psychological therapy trials can be assessed using a simple framework. J Clin Epidemiol. 2020;124:8-15. doi:10.1016/j.jclinepi.2020.04.005

30. Williams VS, Morlock RJ, Feltner D. Psychometric evaluation of a visual analog scale for the assessment of anxiety. Health Qual Life Outcomes. 2010;8:57. doi:10.1186/1477-7525-8-57

31. Hoppitt L, Mathews A, Yiend J, Mackintosh B. Cognitive bias modification: the critical role of active training in modifying emotional responses. Behav Ther. 2010;41(1):73-81. doi:10.1016/j.beth.2009.01.002

32. Smith L, Leung WG, Crane B, Parkinson B, Toulopoulou T, Yiend J. Bilingual comparison of Mandarin and English cognitive bias tasks. Behav Res Methods. 2018;50(1):302-12. doi:10.3758/s13428-017-0871-0

33. Rude SS, Wenzlaff RM, Gibbs B, Vane J, Whitney T. Negative processing biases predict subsequent depressive symptoms. Cognition and Emotion. 2002;16(3):423-40. https://doi.org/10.1080/02699930143000554

34. Rude SS, Valdez CR, Odom S, Ebrahimi A. Negative cognitive biases predict subsequent depression. Cognitive Therapy and Research. 2003;27:415-29. https://doi.org/10.1023/A:1025472413805.

35. Smith L, Leung WG, Crane B., Parkinson B, Toulopoulou T, Yiend J. (2017) Bilingual comparison of Mandarin and English cognitive bias tasks. Behavior Research Methods, 50 (1), 302-312, doi: 10.3758/s13428-017-0871-0.

36. Freeman D, Loe B, Kingdon D, Startup H, Molodynski A, Rosebrock L, Brown P, Sheaves B, Waite F, Bird J. (2021). The revised Green et al., Paranoid Thoughts Scale (R-GPTS): Psychometric properties, severity ranges, and clinical cut-offs. Psychol Med. 2021;51(2):244-53. doi:10.1017/S0033291719003155

37. Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia Bulletin, 13, 261–276. doi:10.1093/schbul/13.2.261

38. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67(6):361-70. doi:10.1111/j.1600-0447.1983.tb09716.x

39. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract. 2008;58(546):32-6. doi:10.3399/bjgp08X263794

40. Spurgeon L, James G, Sackley C. The Hospital Anxiety and Depression Scale: a pilot study to examine its latent structure and the link between psychological state and symptom severity in transient ischemic attack patients. Psychol Health Med. 2016;21(5):632-8. doi:10.1080/13548506.2015.1074711

41. Freeman D, Bird JC, Loe BS, et al. The Dunn Worry Questionnaire and the Paranoia Worries Questionnaire: new assessments of worry. Psychol Med. 2020;50(5):771-80. doi:10.1017/S0033291719000588

42. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes. 2007;5:63. doi:10.1186/1477-7525-5-63

43. Neil ST, Kilbride M, Pitt L, Nothard S, Welford M, Sellwood W, Morrison AP. The questionnaire about the process of recovery (QPR): A measurement tool developed in collaboration with service users: Corrigendum. Psychosis: Psychological, Social and Integrative Approaches. 2010;2(1):88-91. https://doi.org/10.1080/17522430903553495

44. Law H, Neil ST, Dunn G, Morrison AP. Psychometric properties of the questionnaire about the process of recovery (QPR). Schizophr Res. 2014;156(2-3):184-9. doi:10.1016/j.schres.2014.04.011

45. Keetharuth AD, Brazier J, Connell J, Bjorner JB, Carlton J, Taylor Buck E, Ricketts T, McKendrick K, Browne J, Croudace T, Barkham M. Recovering Quality of Life

(ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. Br J Psychiatry. 2018;212(1):42-9. doi:10.1192/bjp.2017.10

46. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull. 2014;40(6):1244-53. doi:10.1093/schbul/sbu033

47. Aitken RC. Measurement of feelings using visual analogue scales. Proc R Soc Med. 1969;62(10):989–93.

48. Savulich G, Shergill SS, Yiend J. Interpretation Biases in Clinical Paranoia. Clinical Psychological Science. 2017;5(6):985-1000. doi:10.1177/2167702617718180

49. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry. 2000;31(2):73-86. doi:10.1016/s0005-7916(00)00012-4

50. First MB, Williams JBW, Karg RS, Spitzer RL. User's Guide for the Structured Clinical Interview for DSM-5 Disorders, Research Version (SCID-5-RV). Arlington, VA: American Psychiatric Association; 2015.

51. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ. 2002;325(7365):652-4. doi:10.1136/bmj.325.7365.652

52. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. J Biopharm Stat. 2001;11(1-2):9-21. doi:10.1081/BIP-100104194

53. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association. 1977;72(358):320-38. https://doi.org/10.2307/2286796

54. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med. 2005;24(7):993-1007. doi:10.1002/sim.1981

55. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons Inc.; 1987. http://dx.doi.org/10.1002/9780470316696

56. Downey RG, King C. Missing data in Likert ratings: A comparison of replacement methods. J Gen Psychol. 1998;125(2):175-91. doi:10.1080/00221309809595542

57. Roth PL, Switzer FS, Switzer DM. Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. Organizational Research Methods. 1999;2(3):211-32. doi:10.1177/109442819923001

58. Ritz C, Streibig JC. Bioassay analysis using R. Journal of Statistical Software. 2005;12(5),1–22. https://doi.org/10.18637/jss.v012.i05

59. Plummer M, Stukalov A, Denwood M. Bayesian graphical models using MCMC, R package version 4-11. https://cran.r-project.org/web/packages/rjags/rjags.pdf. Published October 13, 2021. Accessed November 4, 2021.

60. Ritz C, Streibig JC. Analysis of dose-response curves, R package version 3.0-1. https://cran.r-project.org/web/packages/drc/drc.pdf, Published August 30, 2016. Accessed November 4, 2021.

61. European Commission. Guidelines on Medical Devices (MEDDEV 2.7/3, rev. 3). https://www.medical-device-regulation.eu/wp-content/uploads/2019/05/2_7_3_rev_5_2015_en.pdf. Published May 2015. Accessed November 4, 2021.

62. Mc Glanaghy E, Jackson JL, Morris P, Prentice W, Dougall N, Hutton P. Discerning the adverse effects of psychological therapy: Consensus between experts by experience and therapists [published online ahead of print, 2021 Jul 14]. Clin Psychol Psychother. 2021;10.1002/cpp.2648. doi:10.1002/cpp.2648

63. Freeman D, Garety P. Advances in understanding and treating persecutory delusions: a review. Soc Psychiatry Psychiatr Epidemiol. 2014;49(8):1179-89. doi:10.1007/s00127-014-0928-7

64. Mathews A. Effects of modifying the interpretation of emotional ambiguity. J Cogn Psychol. 2012;24(1):92-105. https://doi.org/10.1080/20445911.2011.584527

65. Savulich G, Shergill S, Yiend J. Biased cognition in psychosis. J Exp Psychopathol. 2012;3:514-36. doi:10.5127/jep.016711

66. May CR, Finch T, Ballini L, MacFarlane A, Mair F, Murray E, Treweek S, Rapley T. Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit. BMC Health Serv Res. 2011;11:245. doi:10.1186/1472-6963-11-245

67. Freeman D. Improving cognitive treatments for delusions. Schizophr Res. 2011;132(2-3):135-9. doi:10.1016/j.schres.2011.08.012

68. Haddock G, Eisner E, Boone C, Davies G, Coogan C, Barrowclough C. An investigation of the implementation of NICE-recommended CBT interventions for people with schizophrenia. J Ment Health. 2014;23(4):162-5. doi:10.3109/09638237.2013.869571

69. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337:a1655. doi:10.1136/bmj.a1655

70. Freeman D, Garety PA, Bebbington PE, Smith B, Rollinson R, Fowler D, Kuipers E, Ray K, Dunn G. Psychological investigation of the structure of paranoia in a non-clinical population. Br J Psychiatry. 2005 May;186:427-35. doi: 10.1192/bjp.186.5.427. PMID: 15863749.

71. Mathews A, Ridgeway V, Cook E, Yiend J. Inducing a benign interpretational bias reduces trait anxiety. J Behav Ther Exp Psychiatry. 2007 Jun;38(2):225-36. doi: 10.1016/j.jbtep.2006.10.011. Epub 2006 Dec 1. PMID: 17141179.

72. Keerie C, Tuck C, Milne G, Eldridge S, Wright N, Lewis SC. Data sharing in clinical trials - practical guidance on anonymising trial datasets. Trials. 2018 Jan 10;19(1):25. doi: 10.1186/s13063-017-2382-9. PMID: 29321053; PMCID: PMC5763739.

73. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. Trials. 2010 Jan 29;11:9. doi: 10.1186/1745-6215-11-9. PMID: 20113465; PMCID: PMC2825513.

74. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. BMJ. 2015 Mar 20;350:h1139. doi: 10.1136/bmj.h1139. PMID: 25794882; PMCID: PMC4707567.

75. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C et al. Consensus statement on defining and measuring negative effects of internet interventions. Internet Interventions. 2014;2014(1):12-19. 1. https://doi.org/10.1016/j.invent.2014.02.001

76. Gullickson KM, Hadjistavropoulos HD, Dear BF, Titov N. Negative effects associated with internet-delivered cognitive behaviour therapy: An analysis of client emails. Internet Interv. 2019 Sep 4;18:100278. doi: 10.1016/j.invent.2019.100278. PMID: 31890627; PMCID: PMC6926340.

77. Kotov RI, Bellman SB, & Watson DB. Multidimensional Iowa Suggestibility Scale (MISS) Brief Manual. 2004.

78. Hettige NC, Bani-Fatemi A, Kennedy JL, De Luca V. Assessing the risk for suicide in schizophrenia according to migration, ethnicity and geographical ancestry. BMC Psychiatry. 2017 Feb 9;17(1):63. doi: 10.1186/s12888-016-1180-3. PMID: 28183281; PMCID: PMC5301397.

79. Imran, Jonathan B. MD; Richmond, Robyn E. MD; Madni, Tarik D. MD; Roaten, Kimberly PhD; Clark, Audra T. MD; Huang, Emily Y.; Mokdad, Ali A. MD; Taveras, Luis R. MD; Abdelfattah, Kareem R. MD; Cripps, Michael W. MD, MSCS; Eastman, Alexander L. MD, MPH Determining suicide risk in trauma patients using a universal screening program, Journal of Trauma and Acute Care Surgery: July 2018 - Volume 85 - Issue 1 - p 182-186. doi: 10.1097/TA.0000000000001899.

80. Rozental A, Kottorp A, Forsström D, Månsson K, Boettcher J, Andersson G, Furmark T, Carlbring P. (2019). The Negative Effects Questionnaire: Psychometric properties of an instrument for assessing adverse and unwanted events in psychological treatments. Behavioural and Cognitive Psychotherapy, 47(5), 559-572. doi:10.1017/S1352465819000018

81. International Council for Harmonisation (2019). ICH E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline om statistical

principles for clinical trials E9(R) (https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf). 2019.

82. Kahan, B.C., Morris, T.P., White, I.R. *et al.* Estimands in published protocols of randomised trials: urgent improvement needed. *Trials* **22,** 686 (2021). https://doi.org/10.1186/s13063-021-05644-4

83. Carpenter, J. and Smuk, M. (2021) Missing data: A statistical framework for practice. Biometrical Journal 63, 915-947

84. White, I. R. & Thompson, S. G. (2005). Adjusting for partially missing baseline measurements in randomized trials. Stat Med 24, 993-1007.

85. Senn S. Contribution to the discussion of "'A critical evaluation of the current p-value controversy"'. Biom J. 2017 Sep;59(5):892-894. doi: 10.1002/bimj.201700032. Epub 2017 Apr 27. PMID: 28452192.

86. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23(6):1352-71. doi: 10.1080/10543406.2013.834911. PMID: 24138436.

87. Carpenter, J. R., & Kenward, M. G. (2013). Multiple imputation and its application (1st ed.). Chichester: Wiley. https://doi.org/10.1002/9781119942283

APPENDICES
Appendix 1 – Version control document

VER No.	DATE (DD/MM/YY)	DATE APPROVED BY HTA	DATE APPROVED BY DMC/TSC	DATE APPROVED BY R&D	DATE APPROVED BY REC/HRA	DATE APPROVED BY MHRA	DATE IMPLEMENTED	COMMENTS e.g. reason for change, stage of study, status (draft or track changes visible), date sent to co-Is or participating sites, acknowledgement of receipt, etc
1.0	19/11/22							Original submission
1.01	09/01/22				14/01/22			Changes made were at the request of the REC as part of our response to provisional opinion
2.0	07/02/22					N/A - objection		Changes made were at the request of the MHRA as part of our response to a request for more information
3.0	27/04/22							Revisions made as per MHRA 'Response to Objection' document
3.1	29/06/22							Revisions made as per MHRA 'Request for more information'

Key
version 0.x = prior to ethics submission
version 1.x = at and during initial ethics submission & approval process
version 2.x = subsequent to ethics approval (e.g. amendments)

version 3.x = MHRA-related amendments

Notes

- Minor amendments (i.e. those that only require notification to ethics, not full review and approval) increase the number after the decimal point by 1
- Substantial amendments ((i.e. those requiring ethics review and approval) increase the number before the decimal point by 1, with the number after the decimal point reverting to 0. (e.g., V2.1 to V3.0).

Appendix 2 – Data Management Plan

0. Proposal name

STOP- Successful Treatment of Paranoia: Replacing harmful paranoid thoughts with better alternatives

1. Description of the data

1.1 Type of study

A Phase II evaluation of a digital mental health therapeutic, the 'STOP' android mobile app. The evaluation comprises a 3 arm double blind randomised controlled design to compare treatment as usual (TAU) plus 6 sessions of STOP; TAU plus 12 sessions of STOP and TAU plus active text reading control. Analysis involves i) a 0, 6, 12, 18 & 24 week follow up of treatment efficacy and ii) a retrospective dose-response modelling
The study will also create the digital therapeutic, STOP, during the first year. While the app itself and the sessional content it delivers (480 scenarios) will be a product of the study, these are classed as arising knowledge with more complex intellectual property rights as detailed in sections 7 and 8 of the application. They are not therefore referred to in this document.
App useage and associated metrics are classed as relevant data and included in this document.

1.2 Types of data

Quantitative: baseline; 0, 6, 12, 18 & 24 week follow ups; pre and post intervention sessions mood trackers. Data generated are from: i) Evaluation: semi-structured clinical interviews, self-report questionnaires (mood, personality, symptom, recovery measures), experimental tasks;
ii) Mobile app: metrics such as useage data and other analytics.

1.3 Format and scale of the data

All the data generated in the evaluation are amenable to sharing and have long-term validity. We expect two databases hosted by the King's CTU, one for dose finding and one for following up and including around 100 variables each. These will be available in both STATA and SPSS formats. Each database will include around 273 cases (patients).

2. Data collection / generation

2.1 Methodologies for data collection / generation

Data will be collected in line with CONSORT guidelines (1). Data will be generated either within the app by direct entry of participant responses, or by researchers collecting responses to individual measures from participants. The trial will be registered with ISRCTN prior to its start date (or as required within 12 months). The trial protocol will be added to the ISRCTN registry within 12 months of the trial starting. The study protocol will be published in a suitable journal (e.g. Trials) with sufficient detail to facilitate replication by other teams. 1. *Kenneth F Schulz, et. al., for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332 doi: 10.1136/bmj.c332*

2.2 Data quality and standards

We use double data entry into SPSS for all manually entered data. The two files are then cross-checked with an automated comparison programme and inconsistencies investigated and corrected. Where possible data collection and entry is automated. Final datasets are checked for quality by outputting key descriptives for all variables and ensuring that values and ranges are within acceptable & plausible parameters. This simple but very effective step ensures raw score errors due to typos etc (e.g. an age of 150, a questionnaire score out of possible range) are easily spotted and corrected.
We have been using the standards and processes set out by the UK Data Archive (e.g. Managing and Sharing Data: Best Practice for Researchers) for managing and sharing data, including their guidance to documenting data and metadata. This ensures our data is of high quality and standards.

3. Data management, documentation and curation

3.1 Managing, storing and curating data.

Data will be stored only on institutional servers at IOPPN and Bath. These are routinely backed up and firewall protected. During the lifetime of the trial, datasets will be managed by the CTU and statisticians. We will follow King's CTU guidelines for data management, ensure all researchers complete the CTU online training package: http://ctu.co.uk/training/ and comply with the King's Data Management Policy: https://www.kcl.ac.uk/governancezone/research/research-data-management-policy. Researchers will also complete the MRC's research data and confidentiality e-learning module. https://byglearning.co.uk/mrcrsc-lms/course/index.php?categoryid=1. King's also provides a dedicated resource to ensure compliance with GDPR, which includes a requirement to register all studies, and a toolkit to ensure compliance: https://internal.kcl.ac.uk/about/secretariat/business-assurance/compliance/gdpr/GDPR-Toolkit-v1.0.pdf

3.2 Metadata standards and data documentation

To ensure our data is transparent and accessible to others we routinely archive datasets with UK Data Archive and will do so for this project. This includes a range of metadata requirements following prescribed formats as follows:

i. providing a data codebook to clearly explain all variables
ii. providing clear and accessible data labels and value labels (e.g. '1' labelled as male; '2' labelled as female)
iii. completing a top-level template giving full details of the provenance and context for the dataset, including links to relevant publications, the institutional sources of the data (here KLC, Bath) and the funding sources that supported it (here MRC).
iv. providing 'read-me' text file to describe how the data might be accessed, used and interpreted and any other idiosyncratic information.

3.3 Data preservation strategy and standards

King's provides a Records and Data Retention Schedule https://www.kcl.ac.uk/aboutkings/orgstructure/ps/audit/records/retention/rdrs-v7.pdf
Under this schedule the required retention period for clinical trials data involving human participants is 20 years after project completion, in line with MCR and King's guidelines. The responsible person is the Principal Investigator and the data is classed as highly restricted, meaning access is limited to a strictly defined group of university users.
The data generated by this study will be deposited with UK Data Archive, via the ReShare process, see https://www.ukdataservice.ac.uk/deposit-data/how-to.aspx. The PI is very familiar with this system, have used it twice before for large grant funded datasets.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards

Compliance and regulatory requirements relevant to our study include: ISO 27001, ISO 13485, ISO 9001, GDPR and NHS Toolkit. See supporting information/ case for support.

4.2 Main risks to data security

Researchers will complete the MRC's research data and confidentiality e-learning module. https://byglearning.co.uk/mrcrsc-lms/course/index.php?categoryid=1
We will adhere to the MRC endorsed principles outlined for sharing clinical trials data outlined here in the publication 'Good Practice Principles for Sharing Individual Participant Data from Publicly Funded Clinical Trials.' Tudur Smith C, Hopkins C, Sydes M, Woolfall K, Clarke M, Murray G, Williamson P. April 2015. http://www.methodologyhubs.mrc.ac.uk/files/7114/3682/3831/Datasharingguidance2015.pdf

5. Data sharing and access

The quantitative data generated by this study will be deposited with UK Data Archive, via the ReShare process, see https://www.ukdataservice.ac.uk/deposit-data/how-to.aspx. The PI is very familiar with this system, have used it twice before for large grant funded datasets.

5.1 Suitability for sharing

The data will be suitable for sharing after having been fully anonymised to reduce the possibility of re-identification. Participants will be assigned a unique ID within the datasets and datasets will not include any direct or indirect identifiers. We will include a specific item in our informed consent procedures allowing participants to indicate whether they consent to anonymous data sharing via a national repository. The data deposit will be limited to those cases giving consent.

5.2 Discovery by potential users of the research data

Our data will be discoverable via the UK Data Archive. To enhance the awareness of the data and access to it we will include links to the archive data deposit as follows: in all study publications; on the study website; on the MRC Gateway; on Open Science Framework.

We will prepare a data sharing policy, in line with the data sharing principles of the referenced MRC policy which will be made available on the study website and UK Data Archive.

5.3 Governance of access

We will ensure that the criteria and process governing access is transparent and in line with MCR policy. It will include independent oversight. Details will be made available on the study website and with the data deposit. This will vary according to the lifecycle stage of the study (active trial recruitment phase, PI led analysis, archiving stage).

As we anticipate only occasional requests for data access we will follow MRC's suggested Model 2 for access decisions. In this model the PI is responsible for access decisions but draws on the study team and/ or an independent advisor. All decisions are documented and the advisor periodically reviews these and approves the study's access policy and procedures.

As indicated above, UK Data Archive will be used.

5.4 The study team's exclusive use of the data

Within the UK Data Archive system depositors can set an embargo on open access and restrict the type of users who can access the data. For this study we will apply a 12-month embargo to access from the end of the trial to allow the main outputs of the study to be accepted in the peer reviewed literature. Should the main findings be published sooner the embargo period will be correspondingly shortened.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

As indicated above specific consent will be sought from participants for data sharing. The data deposit will be limited to those cases giving consent.

We will set access permissions within UK Data Archive, to allow for the implementation of the governance policy outlined above: i.e. new users would be asked to provide a written request for access with reference to MRC policy on research sharing which would then be considered according to the criteria and process outlined in the study policy on data access.

5.6 Regulation of responsibilities of users

New users would be asked to provide a written request for access with reference to MRC policy on research sharing which would then be considered according to the criteria and process outlined in the study policy on data access.

We will ensure that a data sharing agreement is written and signed by both parties as outlined in MRC guidance section 7. It will cover all the items listed in that section such as: new purpose for which the data will be used; conditions of use; appropriate acknowledgement of original funding source etc.

6. Responsibilities
• *study-wide data management* Daniel Stahl (statistician co-applicant), King's Clinical Trials Unit (director, Caroline Murphy) • *metadata creation* Project Manager, researchers, junior statistician all under guidance from the PI • *data security* Daniel Stahl (statistician co-applicant), King's Clinical Trials Unit (director, Caroline Murphy) • *quality assurance of data.* Andrew Webb, Director of Research Assurance

7. Relevant institutional, departmental or study policies on data sharing and data security	
Policy	**URL or Reference**
Data Management Policy & Procedures	https://www.kcl.ac.uk/governancezone/research/research-data-management-policy
Data Security Policy	https://www.kcl.ac.uk/governancezone/GovernanceLegal/Data-Protection-Policy
Data Sharing Policy	*King's endorses the MRC's data sharing policy: study policy of sharing research data. See also:* https://www.kcl.ac.uk/governancezone/InformationPolicies/Data-Governance-Policy
Institutional Information Policy	https://www.kcl.ac.uk/governancezone/informationpolicies/data-governance-policy
Other: IP and commercialisation policy	https://www.kcl.ac.uk/governancezone/research/intellectual-property-commercial-exploitation-and-financial-benefit
Other: research publication policy Other: Data Breach policy	https://www.kcl.ac.uk/governancezone/research/research-publications-policy https://www.kcl.ac.uk/governancezone/governancelegal/data-breach-procedure

8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their **telephone & email contact details**
Professor Jenny Yiend

Appendix 3 – Qualitative Study Research Protocol

DClinPsy Research Project Proposal Form

Name of trainee: Laura Eid
Project Title: A qualitative study on the acceptability of STOP (successful treatment of paranoia), a digital health intervention
1st Supervisor: Dr. Jenny Yiend
2nd Supervisor: Dr. Pamela Jacobsen
Operational Directorate: Department of psychosis studies at King's College London

Aims and objectives

The aim of this qualitative study is to evaluate the acceptability of an interpretation bias modification digital intervention in individuals with paranoia. The study will be informed by NICE Digital Health Treatments (DHTs) Evidence Framework requirements. It will be used to gather evidence to investigate if the digital intervention STOP (successful treatment of paranoia) is in line with NICE evidence standards for DHTs. In other words, it will explore the user experience of STOP and whether clinicians would prescribe it as a potential treatment for paranoia. Both services users and clinicians will have the opportunity to share their views and opinions of STOP. Questions will revolve around acceptability of the application and barriers of the intervention as well as practicality and views around the use of a digital intervention as opposed to a non-digital intervention. This study will therefore help improve future DHTs.

Literature review

Psychosis is amongst the most debilitating mental health conditions with a lifetime prevalence of 3.5% (1). Although presentation vary between individuals, it has been shown that persecutory delusions are the most frequent type of delusions (2). As they are associated with a high level of distress (3) and are negatively correlated with subjective wellbeing and quality of life (4), persecutory delusions have been an important focus in research. By studying the overlap between distressing experiences in anxiety disorders and persecutory delusions (5, 6), researchers have concluded that an important maintaining mechanism for persecutory delusions is an interpretation bias similar to that present in individuals with anxiety (7) and psychosis (8). Such interpretation bias, which is defined as a tendency to interpret ambiguous information in a threatening or negative way (9), will be the focus of the current study.

Cognitive bias modification for interpretation (CBM-I) is a manipulation technique aimed to study the causal link between interpretation bias and paranoia. CBM-I shows that it is possible to change bias through training (10, 11). Such programs aim to change an individuals' unhealthy cognitive patterns of negatively interpreting ambiguity by training them in thinking about alternative, more neutral or positive interpretations. STOP (successful treatment of paranoia) is a CBM-I digital phone application that presents the individual with a range of different ambiguous scenarios which could be completed with a positive, negative, or neutral ending. The aim is that, by repeatedly exposing the participants to positive endings, they will learn to generate positive resolutions to those ambiguous situations. A recent study has shown that CBM-I leads to a decrease in repetitive negative thinking (12). Similarly, the induction of positive interpretation through CBM-I has been shown to reduce trait anxiety (13, 14) and the induction of benign interpretations reduce intrusive thoughts in worriers (15). However, very few studies have studied the effect of CBM on paranoid psychosis, which is why in 2017, Yiend and colleagues designed a pilot randomized controlled trial (RCT) testing the hypothesis that CBM-I could help reduce interpretation bias in individuals with persecutory delusions (16). The 2017 pilot RCT was exploring the effects of a desktop application of CBM-I, called CBM-pa (CBM for paranoia) on individuals with paranoia.

As the current study explores the views and opinions of a digital intervention, it is important to know that such type of interventions have become increasingly common with time and some have proven to be

effective (17). Zhang and colleagues have studied the effectiveness of web-based intervention and have found that it could help reduce biases in anxiety disorders and depression (18). In a systematic review, Zhang discussed ways to gamify interventions and make them more interactive, such as adding a leader board score, time pressure, digital rewards or creating a 3D environment (19). Such methods could increase the platforms attractiveness and adherence to the intervention and should be kept in mind when designing and improving DHTs, including iterations of the STOP intervention. For example, comments from a qualitative study by Leung et al. in 2019 (20), evaluating the acceptability of the previously mentioned DHT CBM-pa, (16) suggested the need to improve the audio and visual displays of the tasks to increase the participant's attention to the task. Changes were therefore made when designing the digital app STOP, which this current study is based on. CBM-pa and STOP differ in a multitude of ways, meaning that the previous qualitative evaluation of CBM-pa cannot be assumed to apply to STOP. Foremost, CBM-pa is a desktop version of CBM-I whereas STOP is a mobile phone application which means that the users' experience of each version differs greatly. With CBM-pa, the participant completes 6 CBM-I sessions in a controlled laboratory setting whereas STOP is a 12-session product that can be used independently outside of the lab setting. The user's experience of attending IOPPN and sitting in an academic office space in front of a computer screen is fundamentally different from the experience of downloading an app onto a mobile phone and using this to receive the therapy. A further key difference is the audio-visual enhancements and interactive manner of presentation of the core therapeutic material within the app environment. This means that while the core therapeutic content (the text) is the same for the first 6 of the 12 sessions, the way this will be experienced by the user is entirely different. Further, unlike CBM-pa, STOP will permit the user to schedule sessions directly and independently and will send automated reminders and alerts. Finally STOP will be twice as long as CBM-pa in its 12-session version. Without a qualitative study of the sort proposed here it will be impossible to know whether STOP in its current form is a product that would be acceptable to users and clinicians and suitable for use in routine clinical practice. The importance of gathering evidence of this sort for any potential new digital therapeutic has been made explicit in the updated Evidence Standards Framework for Digital Health Technologies published by NICE in 2021, which requires evidence of 'credibility with UK Health and Social Care professionals' and 'acceptability with users' under Tier C.

Additional benefits to the use of mobile app intervention are its potential to contribute to the solution against the long waiting time to access NHS psychological services as well as its low cost compared to in person interventions. Furthermore, it would be considered a low stigma intervention where the service users would have private access to their intervention at any location. It could additionally be seen as empowering for the service user who becomes increasingly involved in their care.

The current study will be a qualitative study which will allow different stakeholders to voice their attitudes, and experiences and to get feedback on a new digital intervention called STOP. This would help improve the DHT, therefore ensuring that its standards are in line with the requirements of the NHS digital framework (21) which state that the DHT is viewed as useful and credible within the UK healthcare system, that experts in the field have been involved in the design, development and testing of the DHT or have informally approved it and that service users are satisfied with the DHT as well as have either been involved in designing, developing or testing it. Furthermore, the collection of quantitative data may help contextualise the qualitative data and enhance our understanding of users' experience. This study will therefore be a start point to the translation of the results from a controlled research settings into real life settings, therefore helping us understand and identify service users view of the application and explore if clinicians could visualise themselves prescribing it to their clients. In that sense, this study will be an important next step on translational research.

Hypotheses

The purpose of the study is to explore different views and opinions of the digital intervention and to identify the most important strengths and weaknesses of the DHT as well as potential improvements. DHTs have received positive feedback over the past few years. In effect, qualitative findings from a 2021 study confirmed that a smartphone CBM-I intervention was viewed optimistically (22). Moreover, the 2019 Leung qualitative study showed that the desktop computer based CBM-I intervention CBM- pa is generally well accepted (20). The latter study was used to identify any potential improvement of the CBM-pa, which were then implemented in the development of the phone application STOP. The current study will evaluate the implementation of that

feedback both from the service user and the clinician's perspectives. It is expected that different participants have different views of STOP based on personal preferences. However, the aim of the study is to focus on a consensus of opinions across participants, which will be achieved by coding and analysing the data using the framework method and reporting descriptive data of the user experience questionnaire (UEQ) (23).

Method and Design

The current study will be based on a 4-year-long randomized controlled trial that will run in parallel and that will be looking at the effect of the digital intervention on individuals with paranoia. That main RCT is based on Yiend et al's pilot study in 2017 (16) and will be investigating the effectiveness of STOP, a mobile application CBM-I (16).

Participants:

One of the main RCT's eligibility criteria is to reach a minimum threshold in the Similarity Rating Task (SRT) score, which is an interpretation bias rating score. Such eligibility criterion is crucial as it ensures that the intervention is targeted at a population that does have a negative cognitive bias regarding interpretation of ambiguous scenarios and would benefit from it. Participants included in the current study will consist of those who narrowly miss meeting this inclusion criteria for the main trial. The table below shows that the only inclusion or exclusion criteria that differ between participants eligible for the main RCT and participants included in the current study is the second item related to paranoid concerns. Although this subgroup of participants will not be eligible for the main RCT, they will be eligible for this intervention in routine clinical care if rolled out in the future. In other words, if STOP becomes a routine clinical practice used to decrease interpretation bias in individuals with paranoia, then the objective measure (SRT) will get replaced by self-report of paranoid concerns and clinical judgements.

Table 1: inclusion and exclusion criteria of both the main RCT (16) and the current study

	Main RCT (16)	Current study
Inclusion criteria	1. Presence of significant persecutory or paranoid symptoms as defined by a score of 3 or above on the Positive And Negative Syndrome Scale (PANSS) item 6 related to suspiciousness/persecution. 2. SRT score of 1 or below 3. Medication stability over the past 3 months and throughout the study 4. Between 18 and 65 years old 5. Capacity to consent	1. Presence of significant persecutory or paranoid symptoms as defined by a score of 3 or above on the Positive And Negative Syndrome Scale (PANSS) item 6 related to suspiciousness/persecution. 2. Self-report paranoid interpretations 3. Medication stability over the past 3 months and throughout the study 4. Between 18 and 65 years old 5. Capacity to consent
Exclusion criteria	1. Severe cognitive impairment or illiteracy 2. Important physical illness 3. Substance misuse 4. Currently (or in the past 3 months) receiving psychological intervention targeting interpretation bias	1. Severe cognitive impairment or illiteracy 2. Important physical illness 3. Substance misuse 4. Currently (or in the past 3 months) receiving psychological intervention targeting interpretation bias

Procedure:

One of the aims of the current study is to understand the experience of different session lengths of the intervention and therefore to sample the experience of individuals from both the 6 and 12 sessions intervention. In that sense, 8 service users will purposely be invited to take the 6 sessions intervention and 8 service users will

be invited to take the 12 sessions intervention on a random basis. A total of 16 service users will therefore be recruited to allow for 4 dropouts. Upon completion of the sessions, they will be invited to complete a user experience questionnaire (UEQ), a validated and reliable scale to measure service users' experience of the intervention (23), and a feedback interview. It is important to note that, should a service user not complete the total length of sessions but does remains in the study, they will still be interviewed, and such qualitative data will be important to report. The service users will be made aware of both the feedback interview and the UEQ in the information sheet to ensure the service users' willingness to participate as well as to manage expectations and get informed consent.

Measures:
16 service users who meet the eligibility criteria presented above will be invited for individual semi-structured interviews, and 4 multidisciplinary teams will be invited to attend 4 different focus groups in order to get the view of staff members from different disciplines such as clinical psychologists, psychiatrists, social worker or nurses. Clinicians eligible include MDTs with clinical experience in psychosis but not directly related to the current project. They will be contacted directly through the current study clinicians' network and will be sent a link describing the digital intervention visually a week prior to each focus group for them to explore it at their convenience. A topic guide for the semi-structured interviews will be created based on a previous qualitative study that looked at the acceptability of cognitive bias modification for paranoia (CBM-pa) in patients with psychosis (20).

The 16 recruited service users will also be asked to complete a standardized measure, the UEQ to evaluate their experience of the digital intervention quantitatively. The use of a standardized quantitative measure will help contextualize the qualitative data and enhance our understanding of service users' experience.

Services users will be compensated 40£ for completing the UEQ and the feedback interview which should last an hour. Clinicians will be eligible to receive payment if requested. They will be asked if they would like to claim a payment of 20£ each for their participation and would have the option to fill a form. In our experience in studies of this nature, clinicians elect not to receive payment

Planned analyses

The qualitative data from the interviews and the focus groups will be coded and analysed using the Framework Method (24). As the purpose of the study is to provide evidence that the current DHT is viewed as useful and acceptable by both clinicians and service users, a combination of deductive and inductive approach may be the most relevant. In other words, the analysis will be based on some predefined codes but would also allow for emergent information, which would explain the use of the framework method. Computer Assisted Qualitative Data Analysis Software (CAQDAS), specifically N-Vivo, will be used to aid the coding process and for applying the analytical framework. The analysis of the quantitative data will be done using SPSS. Descriptive statistics will be reported to indicate the overall experience of the intervention using numerical data.

Feasibility and timeline

Service users will be recruited through the main RCT from the institute of Psychiatry, Psychology and Neuroscience (IOPPN), South London and Maudsley (SLaM) NHS trust and Bath, and more particularly the PICCUP clinic, OASIS, MACLEAN charity and C for C. Moreover, clinicians will be recruited by contacting different SLAM services within the field of psychosis and scheduling meetings with MDTs to conduct the focus groups. In total, based on the 2019 qualitative study (20), 12 services users will be recruited and 4 MDTs. Clinicians' data collection will start as soon as ethics approval is received, ideally in January or February 2022 and service users' recruitment will start as soon as the main RCT start recruitment, ideally at the same time as the clinicians' recruitment in early 2022. Transcriptions, coding and UEQ scoring will start progressively, as data is being collected. Framework method analysis and statistical analysis will be completed once all data is collected, which is aimed to be in early 2023.

COVID-19 compliance

If necessary, both focus groups and semi-structured interviews will be conducted online, and the service users will be provided with the necessary equipment such as a mobile or a computer.

Operational Directorate

Preliminary directorate approval has not been given for the project yet. An ethics application through IRAS will be submitted by Christmas 2021.

Costs and budget

Service users will be compensated 40£ for participating in the feedback interview and completing the user experience questionnaire. The total amount for the budget will therefore be 480£.

References

1. Perälä J, Suvisaari J, Saarni SI, Kuoppasalmi K, Isometsä E, Pirkola S, et al. Lifetime prevalence of psychotic and bipolar I disorders in a general population. Archives of general psychiatry. 2007;64(1):19-28.
2. Cutting J. Principles of Psychopathology Two Worlds, Two Minds, Two Hemispheres. 1997.
3. Freeman D. Persecutory delusions: a cognitive perspective on understanding and treatment. The Lancet Psychiatry. 2016;3(7):685-92.
4. McAuley-Wilkinson C. Exploring the relationship between persecutory ideation and subjective wellbeing: The University of Liverpool (United Kingdom); 2021.
5. Clark DM. Anxiety disorders: Why they persist and how to treat them. Behaviour research and therapy. 1999;37(1):S5.
6. Freeman D, Garety PA. Worry, worry processes and dimensions of delusions: an exploratory investigation of a role for anxiety processes in the maintenance of delusional distress. Behavioural and Cognitive Psychotherapy. 1999;27(1):47-62.
7. Chen J, Short M, Kemps E. Interpretation bias in social anxiety: A systematic review and meta-analysis. Journal of Affective Disorders. 2020.
8. Savulich G, Shergill SS, Yiend J. Interpretation biases in clinical paranoia. Clinical Psychological Science. 2017;5(6):985-1000.
9. Trotta A, Kang J, Stahl D, Yiend J. Interpretation bias in paranoia: a systematic review and meta-analysis. Clinical Psychological Science. 2021;9(1):3-23.
10. Mathews A, Mackintosh B. Induced emotional interpretation bias and anxiety. Journal of abnormal psychology. 2000;109(4):602.
11. Yiend J, Mackintosh B, Mathews A. Enduring consequences of experimentally induced biases in interpretation. Behaviour research and therapy. 2005;43(6):779-97.
12. Hirsch CR, Krahé C, Whyte J, Bridge L, Loizou S, Norton S, et al. Effects of modifying interpretation bias on transdiagnostic repetitive negative thinking. Journal of consulting and clinical psychology. 2020;88(3):226.
13. Mathews A, Ridgeway V, Cook E, Yiend J. Inducing a benign interpretational bias reduces trait anxiety. Journal of behavior therapy and experimental psychiatry. 2007;38(2):225-36.
14. Steinman SA, Teachman BA. Modifying interpretations among individuals high in anxiety sensitivity. Journal of Anxiety Disorders. 2010;24(1):71-8.
15. Hirsch CR, Hayes S, Mathews A. Looking on the bright side: accessing benign meanings reduces worry. Journal of abnormal psychology. 2009;118(1):44.
16. Yiend J, Trotta A, Meek C, Dzafic I, Baldus N, Crane B, et al. Cognitive Bias Modification for paranoia (CBM-pa): study protocol for a randomised controlled trial. Trials. 2017;18(1):1-13.
17. Zhang M, Ying J, Song G, Fung DS, Smith H. Mobile phone cognitive bias modification research platform for substance use disorders: protocol for a feasibility study. JMIR research protocols. 2018;7(6):e153.
18. Zhang M, Ying J, Song G, Fung DS, Smith H. Web-Based cognitive bias modification interventions for psychiatric disorders: Scoping review. JMIR mental health. 2019;6(10):e11841.
19. Zhang MW, Ying JB, Song G, Ho R. A review of gamification approaches in commercial cognitive bias modification gaming applications. Technology and Health Care. 2018;26(6):933-44.
20. Leung C, Fosuaah A, Frerichs J, Heslin M, Kabir T, Lee T, et al. A qualitative study of the acceptability of cognitive bias modification for paranoia (CBM-pa) in patients with psychosis. BMC psychiatry. 2019;19(1):1-13.
21. NICE. Evidence standards framework for digital health technologies. NICE London, UK; 2018.

22. Beard C, Ramadurai R, McHugh RK, Pollak J, Björgvinsson T. HabitWorks: Development of a CBM-I smartphone app to augment and extend acute treatment. Behavior Therapy. 2021;52(2):365-78.
23. Laugwitz B, Held T, Schrepp M, editors. Construction and evaluation of a user experience questionnaire. Symposium of the Austrian HCI and usability engineering group; 2008: Springer.
24. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC medical research methodology. 2013;13(1):1-8
25. Freeman, D, Garety, PA, Bebbington, PE, Smith, B, Rollinson, R, Fowler, D et al. British Journal of Psychiatry. 2005; 186:427-35.