iMAPS-2 Statistical and Health Economics Analysis Plan (SHEAP)

Full/ Long title of the Trial: iMAgery focused therapy for Psychosis (iMAPS-2): A feasibility randomised controlled trial

Short Study title/ Acronym: iMAgery focused therapy for Psychosis (iMAPS-2):

Trial registration number(s): ISRCTN81150786

SHEAP version number with dates: v1.0 30th May 2024

Protocol version number and date: 5.0, 27th Oct 2023

SHEAP revisions:

Protocol version	Updated SAP version number	Section number changed	Description of and reason for change	Date changed	Initials
5.0	1.0		First issue		YS

Roles and responsibility: names, affiliations and roles of SAP contributors

Signatures

Trial Statistician: Yvonne Sylvestre, Research Fellow in Clinical Trials Statistics

Signature: 

Date: 04/Jun/2024

ddmmmyyyy

Name: (please print):

.....Yvonne Sylvestre.....

Lead Statistician: Dr Chris Sutton, Senior Lecturer in Clinical Trials Statistics

Signature: 

Date: .03/Jun/2024

ddmmmyyyy

Name: (please print):

..... Dr. Chris Sutton .....

Trial Steering Committee Chair: Professor Craig Steel, Professor of Clinical Psychology and Consultant Clinical Psychologist

Signature: .....  .....

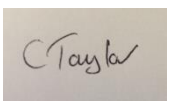
Date: ..4./.6./.2024....

ddmmmyyyy

Name: (please print):

.....

Chief Investigator: Dr Chris Taylor, Consultant Clinical Psychologist & Honorary Senior Lecturer

Signature: 

Date: 30/05/2024

Name: (please print): DR CHRIS TAYLOR

1 CONTENTS

2 ABBREVIATIONS

AE	Adverse Event
CBT	Cognitive Behavioural Therapy
CI	Confidence Interval
CMHT	Community Mental Health Team
CONSORT	Consolidated Standards Of Reporting Trials
EI	Early Intervention Psychosis Team Health Team
ITT	Intention-to-Treat
MID	Minimally Important Difference
NICE	National Institute for Health and Care Excellence
MLE	Maximum Likelihood Estimator
RCT	Randomised Controlled Trial
SAE	Serious Adverse Event
SD	Standard Deviation
SHEAP	Statistical and Health Economics Analysis Plan
TAU	Treatment as Usual

3 INTRODUCTION

3.1 Background and rationale

Schizophrenia is a severe mental health condition, for example, where individuals hear distressing voices other people cannot hear (e.g. auditory hallucinations) and/or have distressing unusual beliefs that others do not share (delusions). They also report negative core beliefs (e.g., “I am vulnerable; I am a failure; others are hostile”) and unwanted intrusive mental images (that are in the “mind’s eye” and other senses) which may be maintaining factors for psychotic symptoms.

One of the best interventions for psychosis is Cognitive Behavioural Therapy (CBT), which is recommended by the UK National Institute for Health and Care Excellence (NICE). Sadly, the first generation of CBT for Psychosis (adapted from CBT for emotional disorders) has a small effect size, and there is a need to refine and improve it. Imagery approaches are almost completely absent from multiple CBT for Psychosis therapy manuals. Empirical studies consistently demonstrate that imagery has a more powerful impact on emotion than verbal cognition. Therefore, we anticipate using an imagery focused approach to target images and schemas will result in a reduction in psychotic symptoms.

We wish to undertake a feasibility randomised controlled trial (RCT) of an imagery focused therapy called iMAPS-2, which targets negative images and negative core beliefs (schemas). Our research aims to improve the current treatments for people with psychosis. We wish to explore a psychological therapy (i.e. iMAPS-2) where the therapist and client specifically work with distressing “mental imagery” (e.g., “pictures in the mind’s eye, sounds in the mind’s ear”), and negative beliefs, which are often reported but rarely treated. Moreover, the project will tell us if we may be able to run a definitive randomised controlled clinical trial. Additionally we will explore proof-of-efficacy on health outcome measures (i.e., Positive and Negative Syndrome Scale (PANSS) Total score, Psychotic Symptom Rating Scales (PSYRATS), Questionnaire about the Process of Recovery (QPR) and Brief Core Schema Scale (BCSS)), but we will also be looking at proof-of-concept (Basic Emotions Scale (BES), Beck Anxiety Inventory (BAI), Calgary Depression Scale (CDS), Warwick Edinburgh Mental Well Being Scale (WEMWBS) and the Personal and Social Performance Scale (PSP)).

3.2 Aim and objectives

The study aims to assess whether it is feasible to conduct a RCT to examine the clinical and cost effectiveness of an imagery focused psychological therapy in psychosis.

Research Questions:

1. What number and percentage of eligible patients/service users consent to the trial (recruitment)?

2. What is the level of engagement with adherence to the iMAPS-2 intervention (therapy sessions attendance measures; therapist fidelity)?

3. What completion and data quality rates can be achieved (data completion and retention of participants)?

4. What estimates of effect sizes (if any) are present (acknowledging this is a feasibility trial)?

5. What are service users' views regarding:

i) acceptability of participating in the trial

ii) the outcomes measures collected, including their acceptability and the ranking of the potential primary outcome measures for the definitive trial, and

iii) acceptability of receiving iMAPS-2 therapy (including adherence to intervention protocol)?

Qualitative component only: not covered in this SHEAP.

6. What is the estimated sample size for a fully powered trial to evaluate the effectiveness of iMAPS-2 (relative to usual care)?

7. What is the range of services used by participants and which are likely to be key cost drivers to consider for the main trial?

8. What is the range of health benefits and are they covered by the EQ-5L health status questionnaire?

4 TRIAL METHODS

4.1 Trial design

This is a feasibility assessor-blind RCT comparing treatment as usual (TAU) plus iMAPS-2 therapy, versus TAU, in a 2:1 allocation ratio.

Participants allocated to the intervention arm received 12 sessions of individualised iMAPS-2 psychological therapy.

iMAPS-2 will also look at the feasibility of establishing the optimum way of evaluating cost-effectiveness.

In addition, iMAPS-2 has a nested qualitative study design of participants' experiences of iMAPS-2 intervention and trial participation, to identify themes relating to the acceptability of this therapy and inform parameters of the definitive trial including therapy protocol, outcome measures, recruitment and retention. However, the qualitative analysis will not be performed by the Trial Statistician or Health Economist and therefore is not described in this document.

4.2 Blinding (masking)

Researchers undertaking assessments were blinded. Any instances of unintentional unblinding were recorded and reported.

The Trial Statistician will remain blinded to treatment allocation until after the SHEAP has been sign off and the analysis of unblinded data commences. It was not possible to blind the participants to allocation group.

4.3 Randomisation

Participants were allocated in a 2:1 ratio to iMAPS-2 intervention plus usual care or usual care alone, using random permuted blocks stratifying by service user/patient under the care of a Community Mental Health Team (CMHT) or an Early Intervention Psychosis Team (EI).

4.4 Sample size

Recruiting 45 participants will allow to estimate the retention rate at end of therapy with a 95% (exact binomial) confidence interval (CI), with width no greater than 25%, assuming retention is at least 80%. It will also suffice for the estimation of the SD, although the sample size (expected minimum 36 participants with outcome data) is towards the lower end of the sample size recommended [1] [2]

4.5 Framework

This is a feasibility trial, within a superiority framework.

4.6 Statistical interim analysis and stopping guidance

Interim Analysis

No interim analysis will be performed.

Guidelines for stopping the trial early

There are no statistical reasons specified for stopping the trial early. The Chair of the TSC reviewed all serious adverse events. The trial would be stopped early to prevent harm, if there were an undue number of related SAE. The TSC also considered any relevant safety data, any recommendations for early closure of the study on this basis would had been

based on their expert opinion, there were no formal statistical criteria specified for stopping the trial early.

4.7 Timing of final analysis

All analyses will be performed at the end of the follow-up period, once all data are either collected or data from any participants with outstanding outcome data is declared missing as agreed by the Chief Investigator and Statisticians.

4.8 Timing of outcome assessments

Outcome assessments will be performed at baseline, 16 weeks (end-of-therapy assessment) and 28 weeks (follow-up assessment) although there is no strict window for outcome assessments.

5 STATISTICAL PRINCIPLES

5.1 Confidence intervals (CI) and level of statistical significance

As this is a feasibility study, the emphasis will be on producing relevant data summaries.. For consideration of potential proof-of-efficacy investigations (i.e., Positive and Negative Syndrome Scale (PANSS) Total score, Psychotic Symptom Rating Scales (PSYRATS), Questionnaire about the Process of Recovery (QPR) and Brief Core Schema Scale (BCSS)), two-sided 100*(1-α)% confidence intervals with α ranging from 0.05 to 0.25 (in steps of 0.05) will be computed [3], thus reporting adjusted point estimates and confidence intervals (CIs), ranging from 75% to 95% confidence, for the effect obtained from the analyses described in section 7.2. This approach is based on a minimally-important difference (MID) between trial arms and is therefore more appropriate than formal hypothesis testing when a study is underpowered.

The MID for the PANSS total score can range between 4 and 21 points [4] [5]. Due to lack of literature to select appropriate MID for the PANSS and the rest primary outcome candidates, we define a promising effect to be around a standard effect size of 0.3 (or more) [ref]. In addition, we will consider an absolute change of 5 points in the Total PANSS scores as a promising effect.

Two-sided 95% confidence intervals will be presented for the estimates of the treatment effects of all other continuous outcomes. No formal testing will be performed and, as a result, p-values will not be presented.

5.2 Adherence and protocol deviations

Adherence

Adherence to the intervention relies upon both therapist and participant factors.

a) Participant adherence

Participant adherence to the iMAPS-2 intervention will be measured by the number of therapy sessions attended by the participants allocated to the intervention arm. An individual will be classed as having attended a session if they attended a part or the whole of a session (defined as a minimum duration of 15 minutes). We will report the mean (SD) number of sessions attended or median (IQR), as appropriate. We will also summarise the number (%) of adherent participants with a 95% exact (binomial) confidence interval (for the %). If a participant attends 5 or more of the 12 sessions [4], then they will be deemed adherent.

b) Therapist fidelity

Therapist fidelity will be assessed by:

i) Rating a sample of audio recordings of therapy sess ions using the Cognitive Therapy Rating Scale Revised (CTS-R) or the Cognitive Therapy Rating Scale for Psychosis (CTS-Psy). Scale scores will be summarised as mean (SD) if they are (approximately) symmetrically distributed or median (IQR) if they are asymmetrically distributed.

ii) Therapists who are part of the research team (approximately 4 therapists) will complete the iMAPS fidelity scale after each therapy session to record the contents of each session in terms of: agenda targets; between-session activities/home tasks; the techniques used. Items within each domain of the fidelity scale will be summarised as follows:

Agenda targets

1. Is there a target Image? (yes/no)
2. Distress re image (0-100)
3. Impact on Functioning (0-100)
4. Is there an Agenda? (at least 1 item included)

We will obtain the frequency (%) of all sessions where both a target image is present and there is an agenda with at least 1 item included. Distress image and Impact on functioning will each be summarised by mean (SD) or median (IQR) as appropriate.

Between-Session Activities(BSA)/Home tasks

1. Last session BSA reviewed as part of agenda? YES/NO
2. Has it been COMPLETED? YES/NO
3. Current session BSA agreed (collaboratively)? YES/NO
4. Last session therapist BSA reviewed as part of agenda? YES/NO
5. Has it been COMPLETED? YES/NO
6. Current Session Therapist BSA set? YES/NO

We will obtain frequencies (%) of all sessions where:

- The previous session home task was reviewed
- The previous session home task was completed
- A new session home task was set

And report them separately for the client (items 1 to 3) and therapist (items 4 to 6).

Technique sessions (Yes, took place in session or NO, did not)

1. Assessment and Engagement
2. Imagery Problem List
3. iMAPS Psychological Formulation (This could be with either the maintenance formulation, longitudinal formulation or both)
4. Psychoeducation regarding power of imagery
5. Normalising the image
6. Safe Place Image
7. Image Suppression
8. Behavioural Experiments
9. Upsetting Memories – Transformation
10. Upsetting Memories – Running image past the worst point
11. Upsetting memories – updating aspects of the image
12. Emotional bridge to past (diagnostic imagery exercise)
13. Imagery rescripting – past events
14. Imagery rescripting – flashforwards
15. Imagery rescripting – discussion of negative beliefs re self and others, schemas
16. Working with night-time imagery - Updating aspects of the image, rescripting new endings
17. Creating Positive Imagery – generating positive images

We will obtain the frequency (%) of all sessions by technique used.

Protocol deviations

We will present a listing of all the protocol deviations occurred during the trial by treatment group with details of the type of deviation recorded. No formal statistical analysis will be undertaken.

5.3 Analysis populations

Feasibility analyses will include all the consented participants to estimate the proportion of eligible consented participants and eligible participants randomised. To summarise the number of iMAPS-2 therapy sessions attended, we shall include all the participants allocated to the intervention group. The proportion of adherent participants (attending at least 3 sessions of the iMAPS-2 therapy) will include all the participants allocated to the intervention group.

Analyses of efficacy will be by intention-to-treat (ITT), where all randomised participants for whom outcomes are available will be analysed in their allocated group regardless of their adherence to therapy.

The safety population will also follow the ITT principle and it will comprise all randomised participants.

6 TRIAL POPULATION

6.1 Screening data

We will produce a CONSORT flow diagram [5]. The upper section of the CONSORT flow chart will include;

- Number of patients referred to the study.
- Number of patients screened.
- Number of potentially eligible patients contacted
- Number of patients recruited (i.e., consented)
- Number of participants recruited but not eligible, along with reasons for exclusions.
- Number of eligible participants.
- Number of eligible participants not recruited, along with reasons for non-recruitment.

6.2 Eligibility

The number of non-eligible participants excluded due to not meeting the inclusion/exclusion criteria will be reported in the CONSORT flow chart as stated in section 6.1.

6.3 Recruitment

The lower section of the CONSORT diagram [5] will be used to describe the course of participants through the trial. This will include the:

- Number of participants recruited
- Number of participants recruited but not randomised, included reasons for non-randomisation.
- Number of participants randomised.
- Number of participants allocated to each arm receiving their allocated intervention, and, for the intervention arm, the average number of sessions received.
- Number of participants not fully completing the trial, along with reasons for withdrawal from treatment, withdrawal from follow-up and loss to follow-up (without formal withdrawal from follow-up).
- Number of participants continuing through the trial and,
- Number participants included in the analyses.

We will present overall monthly figures of participants randomised, that will also be tabulated by randomised group.

6.4 Withdrawal/follow-up

We will produce summaries by randomised group of the:

- Number (%) of participants withdrawing from the allocated intervention, along with reasons (if available).
- Number (%) of participants withdrawing fully from the trial (including all follow-up), along with reasons (if available).
- Number (%) of participants lost to follow-up (excluding those who formally withdrew from the trial).

6.5 Baseline patient characteristics

We will summarise the following demographic characteristics both overall and by randomised group:

- Gender (Male/Female/Non-binary/Other)
- Age (years)
- Chart Diagnosis (ICD-11) Schizophrenia, Schizoaffective disorder, delusional disorder, Schizophreniform disorder, Unspecified (non-affective) psychosis – FEP, other.
- ICD-11 Code

- Illness severity as measured by the positive and negative syndrome scale (PANSS), (Mildly ill/Moderately ill/ Markedly ill/ Severely ill)
- Highest education level (Primary school/secondary school/Further education (e.g. College)/Higher education (e.g. University)
- Employment status (Full Time/Part Time/Retired/Voluntary/Student/ Home duties/ Unemployed)
- Marital status (Single/Married/Living with partner/Civil Partnership/Separated/ Divorced/Widowed)
- Living arrangements. Who does the person live with? (Spouse/Partner only, Spouse/Partner plus children, Spouse/Partner plus other/s (not children), Alone, Children only, Parent/s only, Friend/s only, Supported accommodation/hostel, Other)
- Index of Multiple Deprivation (IMD) Decile (derived from Postcode) (adjacent categories will be merged to form three broader categories: 1-3; 4-7; 8-10)
- Ethnicity (Bangladeshi/Indian/Pakistani/Any other Asian background/African/Caribbean/ Any other Black background/ White & Asian/White & Black African/ White & Black Caribbean/ Any other mixed background/ British/ Irish/Any other White background/ Chinese/Any other ethnic group/ Prefer not to answer)
- Treatments received (i.e. past therapy) and current treatment.
- Service type (Early Intervention Psychosis Team, CMHT, Inpatient)
- Inpatient/Outpatient status and legal status will also be reported
- BAI-rated Anxiety (minimal anxiety (0-7), mild anxiety (8-15), moderate anxiety (16-25), severe anxiety (26-63),).
- CDSS rated depression (minimal or absent (0-6), possible major depressive episode(7-27))

Summary measures for the baseline characteristics of each group and overall will be presented as mean and standard deviation, and median and interquartile range (IQR) for continuous variables, and frequency and percentage for categorical variables.

The categories for some of the variables such as ‘Living arrangements’ or ‘ethnicity’ might need to be merged at the time of analysis given the large number of categories and relatively small sample size.

The presence of traumatic events in each randomised group, as measured by the Trauma and Life Events (TALE) Checklist, will be summarised by frequency and percentage and tabulated by:

- Trauma type (item number): war exposure (1) / attachment-related (2 to 4) / witnessed verbal or physical abuse at home (9) / any interpersonal [bullying and discrimination (5&6), sexual abuse (13&14), emotional abuse (7) physical abuse (8&10), emotional neglect (11), physical neglect (12)] / psychosis-related (15 to 17) / criminal justice (18) / non-interpersonal (19) / other trauma (20).
- Multiple exposure: Repeated events (at least 1 item answered 'yes' to 'more than once') / Multiple trauma types (more than 1 type of events reported)
- Trauma timing: Child (any time <age 16 years) / Adult (any time >age 16 years) / Both (any time <age 16 years AND > age 16 years).

The perceived impact and number of trauma types will be summarised by mean and standard deviation or median and interquartile range (IQR), as appropriate.

7 ANALYSIS

The results of the analyses will adhere to the CONSORT 2010 guidelines and its extension for reporting pilot and feasibility studies [5] [6].

7.1 Calculation of clinical scores

This section describes the scoring and item non-response rules for each tool.

In the absence of tool-specific guidance on handling item non-response, we will seek to impute missing items to complete the tools. We will use a pragmatic approach where if no more than 35% of the items within a (sub)scale are missing, they will be imputed using the mean for all the completed items to provide a valid score for the (sub)scale total, otherwise the (sub)scale will be deemed missing. Total scores will only be available if there are valid (i.e. non-missing) scores for all subscales.

A 35% cut off will allow the scales with a small number of items to have 1 item missing, and the scales with larger number of items a relatively small % of missing items without biasing the results. In psychological outcomes, a range between 20% and 50% is considered acceptable.

Psychosis Measures

Positive and Negative Syndrome Scale (PANSS)

The PANSS is a 30-item therapist-rated scale used to evaluate the symptoms of schizophrenia. All 30 items are rated on a 7-point scale (1 = absent; 2 = minimal; 3 = mild; 4 = moderate; 5 = moderate severe; 6 = severe; 7 = extreme) that will be combined to form five subscales: a) Positive Symptoms (P1 Delusions, G9 Unusual thought content, P3

Hallucinatory behaviour, P6 Suspiciousness and persecution, P5 Grandiosity), b) Negative Symptoms (N2 Emotional withdrawal, N1 Blunted affect, N4 Passive apathetic social withdrawal, N6 Lack of spontaneity, N3 Poor rapport, G7 Motor retardation, G16 Active social avoidance), c) Disorganization often termed Cognitive (P2 Conceptual disorganization, G11 Poor attention, N5 Difficulty in abstract thinking, G13 Disturbance of volition, N7 Stereotyped thinking, G5 Mannerisms/posturing, G15 Preoccupation, G10 Disorientation), d) Affect often termed Depression-Anxiety (G2 Anxiety, G6 Depression, G3 Guilt feelings, G4 Tension, G1 Somatic concern) and e) Resistance or Excitement/Activity (P7 Hostility, G14 Poor impulse control, P4 Excitement, G8 Uncooperativeness) [7].
The items will also be combined to obtain a PANSS Total Score that can also be used to measure and categorise illness severity [8].

Scoring: The five subscales are scored by adding up the corresponding items within each subscale, which will produce subscale scores ranging from 4 to 28, 5 to 35, 7 to 49 and 8 to 56 for the resistance, positive symptoms and affective, negative symptoms and disorganisation subscales respectively, with higher scores indicating more severe symptoms.
The PANSS Total Score is obtained by adding up the 30 items producing a score ranging from 30 to 210, with higher scores indicating more severe symptoms. PANSS baseline total scores between 58 and 74 corresponds to "mildly ill", between 75 and 94 to "moderately ill", between 95 and 115 to "markedly ill" and between 116 and 210 to severely ill. (NB We did not set an “illness” criteria for entry to trial but in relation to PANSS, to be experiencing at least mild hallucinations or delusions symptoms – scoring a 3 on P1 or P3 on PANSS – see inclusion criteria).
An improvement score will also be computed for 16-week and 28-week outcomes using the formula [9]:

$$\frac{(PANSS_{baselinetotalscore} - PANSS_{endpointtotalscore})\ x\ 100}{(PANSS_{baselinetotalscore} - 30)}$$

Missing items: Missing items within subscales will be imputed using a mean of the completed items if no more than 35% items within each subscale are missing; no more than 2 missing item for the negative symptoms and disorganisation subscales and 1 missing item for the positive symptoms, affect and resistance subscales.

Psychotic Symptom Rating Scales (PSYRATS)

The PSYRATS is a 17-item therapist-administered instrument designed to quantify the severity of delusions and hallucinations, each item being rated from 0 (absent) to 4 (severe).

The PSYRATS has 2 subscales: the auditory hallucinations subscale (AHS) consisting of 11 items (i.e., Frequency, Duration, Location, Loudness, Beliefs Re: Origin, Amount of negative content of voices, Degree of negative content, Amount of distress, Intensity of distress, Disruption to life caused by voices and Controllability of voices) and the delusions subscale (DS) consisting of 6 items (i.e., Amount of preoccupation, Duration of preoccupation, Conviction, Amount of Distress, Intensity of Distress and Disruption) [10].

Scoring: Each of the subscales are scored by adding up the items within each subscale, which will yield a score ranging from 0 to 44 for the AHS subscale and from 0 to 24 for the DS, with higher scores indicating higher severity of hallucinations and delusions respectively.

Missing items: Missing items within subscales will be imputed using a mean of the completed items if no more than 35% items within each subscale are missing (i.e., 3 and 2 missing items for the AHS and DS subscales respectively).

Questionnaire about the Process of Recovery (QPR)

The QPR is a 15-item self-reported measure of personal recovery. Each item is scored using a 5-point Likert scale ranging from 0 (disagree strongly) to 4 (agree strongly) [11].

Scoring: A total score ranging from 0 to 60 is obtained by summing all the items, with high scores indicating better progress towards recovery goals.

Missing items: Missing items will be imputed using a mean of the completed items if no more than 35% items (i.e. 5 items) are missing.

Imagery

Mental Imagery in Psychosis Questionnaire/Visual Analogue Scales (MIPQ/VAS)

The MIPQ is an instrument used to measure mental imagery. For each image, the participants will complete 7 items rated on a visual analogue scale ranging from 1=not at all to 10=extremely [12].

Scoring: A total score ranging from 7 to 50 is obtained by summing items 1 to 5, with higher scores indicating more problematic imagery. Items 6 and 7 will be reported separately.

Missing items: Missing items 1 to 5 will be imputed using a mean of the completed items if no more than 1 item is missing (i.e., 35% items).

Psychosis Imagery Questionnaire (PIQ)

This tool is not a secondary outcome for iMAPS-2. Validation of the new measure will be performed after the trial is completed, analysis and results of the validation study will be reported elsewhere.

Schemas

Brief Core Schema Scale (BCSS)

The BCSS aims to provide a theoretically coherent self-report assessment of schemata concerning self and others in psychosis. It consists of 24 items concerning beliefs about the self and others that are assessed on a 4-point Likert scale (Believe it: 1=slightly, 2=moderately, 3=very much and 4=totally). Four domains are obtained: negative-self (items 1 to 6), positive-self (items 7 to 12), negative-others (items 13 to 18) and positive-others (items 19 to 24). The participant is asked to indicate in a dichotomous No/Yes format whether they held each belief. Then, if they held the belief (i.e. answered yes), they are asked to indicate their degree of belief conviction by circling a number from 1 to 4. If they do not hold the belief, a score of zero is assigned [13].

Scoring: Each domain is scored by adding the items within each domain, which will yield a score ranging from 0 to 24 with higher scores indicating higher negative/positive evaluation of self/others.

Missing items: Missing items within domains will be imputed using a mean of the completed items if no more than 35% items within each domain are missing (i.e., 2 missing items).

Trauma

International Trauma Questionnaire (ITQ)

The ITQ is an 18-item self-report measure with each item rated on a 5-point Likert scale ranging from 0 (Not at all) to 4 (Extremely) focusing on the core features of Post-Traumatic Stress Disorder (PTSD) and Complex PTSD (CPTSD). The ITQ has two major subscales with three symptom clusters in each:

a) Post-Traumatic Stress Disorder (PTSD); Re-experiencing, Avoidance AND Sense of threat.
b) Disturbances in self-organization (DSO); Affective dysregulation, Negative self-concept and Disturbances in relationships.

The PTSD and DSO symptoms are accompanied by three items measuring associated functional impairments in the domains of social, occupation, and other important areas of life [14].

Scoring: There are two components of scoring and interpretation for this tool; Categorical scoring for the diagnosis of PTSD and CPTSD, and a dimensional component which measures symptom severity. iMAPS-2 is concerned with the dimensional component only.

The PTSD (items 1 to 6) and DSO (items 10 to 15) subscales range from 0 to 24 and they are computed by summing the items within each subscale, with higher scores indicating more severe symptoms. A total score for CPTSD ranging from 0 to 48 can be obtained by summing the 12 items comprising the PTSD and DSO subscales, with higher scores indicating more severe symptoms.

Missing items: Missing items for the PTSD and DSO subscales will be imputed using a mean of the completed items within each subscale if no more than 35% items are missing (i.e., 2 items per subscale). A total score for CPTSD will only be computed if there are valid scores for PTSD and DSO subscales, after item imputation if necessary.

Basic Emotions Scale (BES)

The BES is a self-report measure comprising 20 items each rated on a 7-point Likert scale ranging from 1 (never) to 7 (very often) that can be combined to form 5 subscales: anger (items 1, 6, 11 and 16), sadness (items 2, 7, 12 and 17), disgust (items 3, 8, 13 and 18), anxiety (items 4, 9, 14 and 19), and happiness (items 5, 10, 15 and 20).

Scoring: Each subscale is obtained by summing their corresponding item scores which yield subscales ranging from 4 to 28 with higher scores indicating higher frequency of emotions experienced [15].

Missing items: Missing items within subscales will be imputed using the mean of the completed items in each subscale, if no more than 35% of the items (i.e.,1 item) within each subscale are missing.

Mental Health and Functioning

Beck Anxiety Inventory (BAI)

The BAI is a 21-item self-report instrument in which patients rate their experience of the severity of each symptom of anxiety. Each items is rated on a 4-point Likert scale with the anchors of 0 (Not at All) to 3 (Severe) [16].

Scoring: A total score, ranging from 0 to 63, is obtained by adding each of the item scores, with higher scores indicating higher levels of anxiety.

Missing items: Missing items will be imputed using a mean of the completed items if no more than 35% items (i.e. 7 items) are missing.

Calgary Depression Scale (CDSS)

The CDSS is a 9-item clinician rated outcome measure that assesses the level of depression in people with schizophrenia. Each item is rated on 3-point Likert-scale (0 = absent; 1 = mild; 2 = moderate; 3 = severe) [17].

Scoring: A total score, ranging from 0 to 27, is obtained by adding each of the item scores, with higher scores indicating higher levels of depression.

Missing items: Missing items will be imputed using a mean of the completed items if no more than 35% items (i.e. 3 items) are missing.

Warwick Edinburgh Mental Well Being Scale (WEMWBS)

WEMWBS is a self-reported instrument measuring subjective well-being focusing entirely on positive aspects of mental health. It is composed of 14-items each rated on 5-point Likert scale (1 = none of the time; 2 = rarely; 3 = some of the time; 4 = often; 5 = all of the time) [18].

Scoring: A total score ranging from 14 to 70 is obtained by summing the item responses of all 14 items, with higher scores indicating greater positive mental wellbeing.

Missing items: Missing items will be imputed using a mean of the completed items if no more than 35% items (i.e. 4 items) are missing.

The Personal and Social Performance Scale (PSP)

The PSP is a therapist-reported instrument evaluating the degree of difficulty a participant exhibits over a 1-month period in four domains: (a) personal and social relationships; (b) socially useful activities, including work and study; (c) self-care; and (d) disturbing and

aggressive behaviours. The patient's degree of severity in the four domains is rated on a six-point scale from absent (which means no problems on this dimension) over mild, manifest, marked, severe to very severe difficulties in the given area.

[19].

Scoring: Out of the ratings on the four domains, a total score on a 100-point scale is computed by the therapist according to manual instructions, with higher scores indicating better personal and social functioning.

Missing items: Due to the nature of the PSP scoring is not feasible to impute missing domains to obtain a total score. If a participant declines to answer part of the interview relating to a particular domain, then the PSS will not be scored.

Working Alliance Inventory Short form (WAI-SR)

This is a rating scale designed to measure the working alliance between therapists and their clients during therapy sessions. The client version is composed of 12 items each rated on a 5-point Likert scale (1 = Seldom; 2 =Sometimes; 3 = Fairly often; 4 = Very often; 5 = Always) on three domains: a) Goal: client and therapist agree on goals of treatment (items 4, 6, 8 and 11), b) Task: whether they agree on how to achieve those goals (items 1, 2, 10 and 12) and c) Bond: whether they will be able to or have already established a personal bond with each other (items 3, 5, 7 and 9).

The therapist version is composed of 10 items also rated on a 1-5 point scale on 3 domains: Goal (items 3, 6 and 8), Task (1, 4 and 10) and Bond subscales (2, 5, 7 and 9).

Scoring: A total score is obtained be adding all 12 items producing a scale ranging from 12 to 60, with higher scores indicating greater alliance in the client version.

For the therapist version a total score is obtained be adding all 10 items producing a scale ranging from 10 to 50 with higher scores indicating greater alliance.

Missing items: Missing items within domains will be imputed using the mean of the completed items in each domain, if no more than 35% of the items (i.e.,1 item) within each domain are missing for the client and therapist versions. A total score will be computed, for the client and therapist versions, if no more than 1 missing item per domain is missing.

Adverse Effects

Clinical Global Impression- Improvement (CGI-I)

The CGI I is a stand-alone measure rating the change from baseline to treatment. The CGI I is rated on a 7-point scale: 1=very much improved since the initiation of treatment; 2=much improved; 3=minimally improved; 4=no change from baseline (the initiation of treatment); 5=minimally worse; 6= much worse; 7=very much worse since the initiation of treatment [20].

Participant and therapist versions will be reported separately.

Clinical Global Impression - Severity (CGI-S) Version

The CGI S is a stand-alone measure rating illness severity. The CGI S is rated on a 7-point scale: 1=normal, not at all ill; 2=borderline mentally ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill patients [20].

Participant and therapist versions will be reported separately.

7.2 Analysis methods

Objective 1. What number and percentage of eligible patients/service users consent to the trial (recruitment)?

Outcomes

- Number of patients referred to the study.
- Number of patients (potential participants) contacted
- Number of patients screened for eligibility.
- Number of patients screened but not eligible, along with reasons for non-eligibility or withdrawal of interest.
- Number of eligible patients.
- Number of eligible patients not recruited, along with reasons for non-recruitment
- Number of participants recruited (consented).
- Source of recruitment
- Number of participants randomised.

Analysis

The number of people screened, eligible, consented and randomised will be summarised and reported in a CONSORT flow diagram (see section 6).

Randomisation rates to assess the progression criteria to conduct a definitive trial will be presented as point estimates with (exact binomial) 95% confidence intervals.

Monthly randomisation rates will also be presented overall and tabulated by randomised group as stated in section 6.3

Progression criteria

Progression criteria for recruitment is:

- Green (progress to full trial). At least 80% of target recruited (n ≥ 36)
- Amber (full trial considered feasible if reasons for poor recruitment identified and can be addressed). Between 40% and <80% of target recruited (n between 18 and 35, inclusive)
- Red (unlikely to progress to full trial. Substantial amendments will be necessary to make a full trial feasible). Below 40% of target recruited (n<18).

Objective 2. What is the level of engagement with adherence to the iMAPS-2 intervention (therapy sessions attendance measures; therapist fidelity)?

Outcome

Participant adherence

- Number of sessions of iMAPS-2 therapy attended for those participants allocated to the intervention arm.

Therapist fidelity:

- The Cognitive Therapy Rating Scale Revised (CTS-R) or
- The Cognitive Therapy Rating Scale for Psychosis (CTS-Psy)
- Therapist checklist.

Analysis

Assessment of participant's adherence and therapist fidelity are fully described in section 5.2.1.

The proportion of participants allocated to the intervention arm, attending at least 5 sessions of the iMAPS-2 therapy will be presented as point estimates with 95% (exact binomial) confidence intervals to assess the progression criteria to conduct a definitive trial.

In addition, we will report the proportion of participants whom have received at least 5 hours (300 minutes) of the iMAPS-2 therapy along with a 95% (exact binomial) CI. We will also summarise, as mean(SD) or median(IQR) as appropriate, the overall hours of iMAPS-2

therapy received, and the hours of iMAPS-2 therapy received by the adherent participants (those who received at least 5 sessions).

Finally, we will plot:

- A histogram showing the distribution of iMAPS-2 therapy sessions attended (whether they are part or whole sessions).
- A histogram showing the distribution of hours of iMAPS-2 therapy sessions attended.

Progression criteria

Progression criteria for participant adherence is:

- Green (progress to full trial). At least 75% of the participants allocated to the intervention arm attend at least 5 sessions of iMAPS-2 therapy.
- Amber (full trial considered feasible if reasons for poor participant's adherence identified and can be addressed). 40% - <75% of the participants allocated to the intervention arm attend at least 5 sessions of iMAPS-2 therapy.
- Red (unlikely to progress to full trial. Substantial amendments will be necessary to make a full trial feasible). If less than 40% of the participants allocated to the intervention arm attend at least 5 sessions of iMAPS-2 therapy.

Objective 3. What completion and data quality rates can be achieved (data completion and retention of participants)?

Outcome

- Number of participants continuing in the trial at 16 weeks.
- Number of participants completing the trial at 28 weeks.
- Questionnaire completion rates (after missing item imputation) at baseline, 16 weeks and 28 weeks for the following instruments:
    - Positive and Negative Syndrome Scale (PANSS)
    - Psychotic Symptom Rating Scales (PSYRATS)
    - Questionnaire about the Process of Recovery (QPR)
    - Mental Imagery in Psychosis Questionnaire/Visual Analogue Scales Visual Analogue Scales (MIPQ)
    - Brief Core Schema Scale (BCSS)
- Number of completed items per questionnaire as listed in the previous bullet point and prior to missing item imputation.

Analysis

The number of participants attending each of the 16- and 28-week trial visits will be summarised and reported in a CONSORT flow diagram (see section 6.3). In addition, retention rates at 16 weeks to assess the progression criteria to conduct a definitive trial, and also at 28 weeks, will be presented as point estimates with 95% (exact binomial) confidence intervals.

We will summarise the frequency (%, with denominator the number attending the corresponding visit) of the completed instruments (after missing item imputation) listed above by time point. We will also report the number of completed items per instrument (prior to missing item imputation) at each time point, and frequency (%) of valid scale (and subscale) scores for each instrument.

If any of the candidate primary outcome measures shows a significant amount of missing item data (more than 5% of partially or fully completed questionnaires), we will present the % of missing data by item to identify problematic items or sets of items.

Progression criteria

Progression Criteria for retention (baseline to primary outcome completion at 16-weeks) is:

- Green (progress to full trial). At least 80% retention:
- Amber (full trial considered feasible if reasons for poor retention identified and can be addressed). 60%-<80% retention
- Red (unlikely to progress to full trial. Substantial amendments will be necessary to make a full trial feasible). Below 60% retention.

Objective 4. What estimates of effect sizes (if any) are present (acknowledging this is a feasibility trial)?

Analysis

We shall analyse each of the candidate primary outcomes (i.e., Positive and Negative Syndrome Scale (PANSS) Total score, Psychotic Symptom Rating Scales (PSYRATS), Questionnaire about the Process of Recovery (QPR) and Brief Core Schema Scale (BCSS), using a separate linear regression model at each time-point to estimate the effect sizes at 16 and 28 weeks.

Each model will be fitted to the clinical outcome measure at the respective assessment time-point (16 or 28 weeks) and will include the fixed-effects for treatment therapy (iMAPS-2+TAU or TAU), care team type (CMHT or EI) and the corresponding baseline outcome score.

Models will be fitted using Maximum Likelihood Estimation (MLE).

Point estimates will be presented as regression coefficients and a range of confidence intervals at different confidence levels (75%, 80%, 85%, 90% & 95%) will be structured and examined as whether they are likely to capture the minimal clinically important difference (MID) for the primary end-point (i.e., 16 weeks). Point estimates with 95% confidence intervals will be presented at 28 weeks.

For this purpose, the MID will, in general, be a SES of 0.3, and so the confidence intervals used for comparison with the MID will likewise be standardised using the overall baseline SD. However, for the PANSS Total Score, the confidence intervals for the effect estimates on the unstandardised scale will also be compared with the proposed MID of 5 points.

These analyses will be used to consider evidence as to potential proof of effectiveness and will also provide estimates of their SD to assist with the estimation of the required sample size for a full effectiveness trial (Objective 6), should that measure be retained for consideration as primary. We will report the estimated standardised effect size (SES)= Effect estimate / SD_{pooled} with their corresponding 95%CIs calculated dividing the limits of the 95%CI of the effect estimate by the SD_{pooled}.

The remaining continuous outcomes listed in section 7.1 will be analysed in the same fashion than the candidate primate outcomes. Each regression model will be fitted to the clinical outcome measure at the respective assessment time-point (16 or 28 weeks) and will include the fixed-effects for treatment therapy (iMAPS-2+TAU or TAU), care team type (CMHT or EI) and the corresponding baseline outcome score. Estimates of the treatment effects will be presented as the mean difference with 95% CIs.

Regression diagnostics

Diagnostic checks to assess the appropriateness of the regression models fitted will be through the use of residual plots. We will plot:

- Histograms and probability plots of the residuals to assess normality
- Scatter plots of residuals against fitted values to assess constant variance and linearity, and to identify potential outliers.

Should the normality assumption be untenable for the continuous outcomes, bootstrap methods will be applied to estimate the confidence intervals of the treatment effects.

Objective 6. What is the estimated sample size for a fully powered trial to evaluate the effectiveness of iMAPS-2 (relative to usual care)?

Outcomes

Positive and Negative Syndrome Scale (PANSS), Psychotic Symptom Rating Scales (PSYRATS), Questionnaire about the Process of Recovery (QPR) and Brief Core Schema Scale (BCSS)

Analysis

The analyses described for Objective 4, together with the findings of the qualitative analysis that will consider the views of the participants on the candidate primary outcome measures, will be used to 'shortlist' a potential primary outcome measure(s). We will also consider the feasibility of running a full effectiveness trial with each of those candidate outcomes remaining after this process, by considering the respective sample sizes required to detect a specified MID to assist with our decision-making. We will discuss our findings and proposals with the Trial Steering Committee to help inform our recommendations for a full effectiveness or evaluation trial.

Sensitivity analyses

Follow-up outcome assessments are performed at 16 weeks (end-of-therapy assessment) and 28 weeks (follow-up assessment), although there is no formal time window to maximise completion of outcome data as stated in section 4.8. From a feasibility perspective to report the frequency (%) of participants completing the assessments within a reasonable timeframe (i.e., within 4 weeks of the scheduled time-points).

If the proportion of scheduled assessments completed more than 4 weeks after the scheduled date for any time-point is ≥10%, we will repeat the efficacy analyses for objective 4 excluding the 'out of window' assessments. This will help to explore the consideration of data collection and practicalities around introducing a 'completion window' for a full trial.

7.3 Subgroup analyses

N/A

7.4 Missing data

As this is a feasibility study we will focus on reporting the amount of missing data (see sections 6.4). We will report the number (%) of participants withdrawing fully from the trial and lost to follow-up by treatment arm. Reasons for missingness may be important and these will be tabulated by treatment arm and documented as far as possible.

We will also tabulate the number of missing items within questionnaires (see Objective 3, section 7.2).

Efficacy analyses of clinical outcomes are adjusted for baseline values. If there are any missing baseline values of the corresponding outcome data, we will use simple mean imputation (across the groups) to avoid exclusion of such participants in the proposed complete-case analysis [21].

7.5 Additional analyses

There are 3 additional analyses planned:

I. We will tabulate the PANSS response rates:
   - Deterioration or no change (Negative or 0% reduction from baseline)
   - >0% but <25% reduction from baseline
   - 25% to <50% reduction from baseline
   - 50% to <75% reduction from baseline
   - 75% to <100% reduction from baseline
   - 100% reduction from baseline

   at 16 and 28 weeks by treatment arm to show how many participants are still symptomatic at the end of treatment and follow up, and how many have responded and to show the overall amount of change.

II. To assess the extent of unblinding, we will report the frequency (%) of recorded instances where outcomes assessors reported to be unblinded, overall and by randomised group.

III. We will summarise the individual items of the brief imagery interview (shortened version of Schulze et al. , 2013) where participants describe the images (from 1 image to 3 images) they experienced in relation to your psychosis:
   1. Image frequency in past week (0 = images have not occurred 1 = images occur once a week; 2 = images occur a few times a week; 3 = images occur once a day; 4 = images occur multiple times a day)
   2. Image frequency in past month 0 = images have not occurred 1 = images occur once a week; 2 = images occur a few times a week; 3 = images occur once a day; 4 = images occur multiple times a day
   3. Image Distress (0-100) where 0 is not distressing at all and 100 is the most distressing it could be.

We will tabulate items 1 and 2 and present frequencies (%) of the endorsed individual items. Images reported and Image distress will be summarised by mean (SD) or median (IQR) as appropriate.

Results will be presented by allocated treatment and time point (i.e., Baseline, 16 weeks and 28 weeks).

IV. The reporting of harms, as described in the next section.

7.6 Harms

The number of safety events (e.g., AEs and SAEs) and number (and proportion) of participants reporting at least one safety event will be summarised by treatment arm. Summary measures will be number (%).

Safety events will also be tabulated by:

- Death by suicide
- Suicide attempt
- Suicidal crisis without attempt (rating 2 on item 8 of Calgary Depression Rating Scale for Schizophrenia CDSS)
- Severe symptoms increase (rating of > 6 on the patient or researcher rated CGI and CGI-I)

They will be reported by both overall and by treatment arm.

Treatment effects will be estimated by the difference in event rates and 95% CI for the differences.

In addition, we will tabulate and present frequencies of the endorsed individual items in the Adverse Effects in Psychotherapy (AEP) scale (29) for those participants allocated to the intervention arm. This is a self-report measure of potential adverse effects completed by the participants at the end of therapy assessment measuring 7 broad categories: worsening difficulties, poor engagement (including low motivation), situational change, no benefit from therapy, stigma, conflict with others (family, care team) and feeling better.

7.7 Health economic analysis

The aim of the health economic analysis is to establish the optimum way of evaluating cost-effectiveness in a full trial. The focus is on informing the data required for a definitive trial, including health status measurement and the range of costs to be included.

Health status (EQ-5D) and Quality-Adjusted Life-Years (QALYs)

To assess the feasibility of collecting the EQ-5D measure in a full trial, we will report and review the completeness of the data, as well as an assessment of whether the measure appears to be reflecting health in the population. EQ-5D index values/utility values will be derived using the approach recommended by NICE, which is currently using the validated mapping function from the existing EQ-5D-3L. In line with current NICE recommendations, the mapping function developed by the Decision Support Unit (DSU) using the 'EEPRU dataset' will be used [22].

The key questions that will be answered are below:

- How complete is the EQ-5D data? This will be assessed by reporting the proportion of participants with partially (e.g., missing a domain) or completely missing EQ-5D responses.
- Are the utilities as expected compared to general population norms for a similar age and gender mix? This will be assessed by comparing the estimating utility values to published population norms [25].
- How do the EQ-5D index values correlate to key measures and demographics? This will be assessed by direction, strength of correlation and statistical significance using a simple exploratory regression analysis.

The described analysis will provide some early conclusions on the suitability of the EQ-5D for future trials and will also give some information on the likely moderators or influencers of total QALYs, to inform data collection in the definitive trial.

Recognising some of the limitations of the EQ-5D, the ReQoL-10 (a newer and more mental health focused measure) has also been collected within the feasibility analysis and will be used to generate alternative utility scores using a published algorithm [24]. The completeness of data will again be assessed to identify whether participants found the measure easier or more challenging to complete in comparison to the EQ-5D. In addition, the generated utility scores (EQ-5D and ReQoL) will be compared to assess whether one measure may be more or less suitable.

Health and social care use

As part of the feasibility trial a health service resource use questionnaire is collecting participant reported health and social care use (hospital, primary, community and social care use). An analysis of the range of services used and ability of participants to report complete service use data will be used to inform a definitive trial.

Data from the resource use questionnaire will be cleaned and costed. Total costs will be presented, and mean costs will be reported for each type of service use, alongside standard

deviation and 95% confidence intervals. The unit costs of NHS and social care services will be derived from national average unit cost data. These unit costs are published annually in the NHS reference costs database, and in the Unit Costs of Health and Social Care document published by the Personal Social Services Research Unit (PSSRU), University of Kent.

The key questions that will be answered are below:

- Was the questionnaire feasible for participants and researchers to complete? This will be assessed with a summary of missing data (by type of service) to assess the level of missingness and whether any sections were difficult to complete. Any unfeasible values or outliers will be discussed with the research team. Discussions will be held with the researchers to get their perspective of the questionnaire. This will focus on whether simplifications can be made and whether the recall period is appropriate.
- Would revisions to the health resource use questionnaire be required for a definitive trial? If so, what are these? Following the findings above, suggestions will be made to revise service use collection forms.

Intervention costs

Costs of providing the iMAPS intervention will be estimated using the staff time to deliver intervention and number of sessions.

Exploratory cost-effectiveness analysis

An exploratory cost-effectiveness acceptability analysis may be conducted. However, this will be limited due to the feasibility stage of the work (in particular the sample size) and it should be highlighted that the key aims of the economic analysis are not to assess cost-effectiveness, rather to aid the design of a definitive trial. If the data are sufficient (i.e., if complete cost and QALY data are available for >75% of participants), an exploratory cost-effectiveness analysis will be performed.

Analysis of the economic data will use an intention-to-treat approach and will consider complete cases only. This analysis will use only the observed data and will provide insight to the result for the group of participants with complete follow up and complete data (evaluable cohort). The proportion of participants for which complete QALY and cost data are available will be summarised. In a definitive trial, imputation techniques would be applied to overcome some of the impact of missing data, however, this exploratory analysis will be kept simple to not overstate the usefulness of the analysis.

Costs and health benefit for the primary analysis will be estimated from baseline to end of follow-up, to estimate the incremental cost-effectiveness of the addition of iMAPS intervention. The primary measure of interest for the economic analysis is the incremental cost-

effectiveness ratio (ICER). This is calculated by dividing the difference in costs (net costs) by the difference in QALYs (net QALYs) between any two interventions. The ICER represents the additional cost associated with an intervention per additional QALY gained:

$$\text{ICER} = \frac{\text{Cost}_{\text{iMAPS intervention plus TAU}} - \text{Cost}_{\text{TAU}}}{\text{QALYs}_{\text{iMAPS intervention plus TAU}} - \text{QALYs}_{\text{TAU}}}$$

Regression analysis will be used to estimate the net costs and QALYs of iMAPS intervention. Key covariates (aligned with the statistical analysis plan) will be included in the regression models to control for baseline factors that may influence QALYs or costs. The covariates for these analyses will be identified in the analysis described above. The estimates of costs and health benefits from the regression analyses will be bootstrapped to simulate 10,000 pairs of incremental cost and QALY outcomes of the intervention. Pairs of net costs and QALYs will be plotted on a cost-effectiveness plane to illustrate the level of uncertainty in the data.

Finally, each of the net QALY estimates from bootstrap simulation results will be revalued by multiplying it by a willingness to pay threshold to estimate the net benefit statistic. The monetary value of simulated QALYs will be varied from £0 to £30,000 to reflect a range of hypothetical willingness to pay thresholds (WTPT).

7.8 Statistical software

All analyses will be performed using STATA/SE (StataCorp, College Station, TX, USA).

8 REFERENCES

[1] S. A. Julious, “Sample size of 12 per group rule ofthumb for a pilot study,” *Pharmaceutical Statistics,* vol. 4, no. 4, p. 287–291, 2005.

[2] J. Sim and M. Lewis, “The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency,” *Journal of Clinical Epidemiology,* vol. 65, no. 3, pp. 301-308, 2012.

[3] E. C. Lee, A. L. Whitehead, R. M. Jacques and S. A. Julious, “The statistical interpretation of pilot trials: should significance thresholds be reconsidered?,” *BMC medical research methodology,* vol. 14, no. 1, p. 41, 2014.

[4] S. Jolley, P. Garety, E. Peters , . M. Fornells-Ambrojo, J. Onwumere, . V. Harris and A. Brab, “Opportunities and challenges in Improving Access to Psychological Therapies for people with Severe Mental Illness (IAPT-SMI): evaluating the first operational year of the South London and Maudsley (SLaM) demonstration site for psychosis,” *Behaviour Research and Therapy,* vol. 64, pp. 24-30, 2015.

[5] K. F. Schulz, D. G. Altman and D. Moher, “CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials,” *BMJ,* p. 340:c332 , 2010.

[6] S. M. Eldridge, C. L. Chan, M. J. Campbell, C. M. Bond, S. Hopewell, L. Thabane and G. . A. Lancaster, “CONSORT 2010 statement: extension to randomised pilot and feasibility trials,” *BMJ,* p. 355:i5239, 2016.

[7] A. Shafer and F. Dazzi, “Meta-analysis of the positive and Negative Syndrome Scale (PANSS) factor structure,” *Journal of Psychiatric Research,* vol. 115, pp. 113-120, 2019.

[8] M. van der Gaag, T. Hoffman, M. Remijsen, R. Hijman, L. de Haan, B. van Meijel, P. N. van Harten, L. Valmaggia, M. de Hert, A. Cuijpers and D. Wiersma, “The five-factor model of the Positive and Negative Syndrome Scale II: A ten-fold cross-validation of a revised model,” *Schizophrenia Research,* vol. 85, no. 1-3, pp. 280-7, 2005.

[9] S. Leucht , J. Davis , R. Engel , W. Kissling and J. Kane , “Definitions of response and remission inschizophrenia: recommendations for theiruse and their presentation,” *Acta Psychiatrica Scandinavica,* vol. 119, no. (Suppl. 438), p. 7–14, 2009.

[10] G. Haddock, J. McCarron, N. Tarrier and E. B. Faragher, “Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS),” *Psychological Medicine,* vol. 29, no. 4, pp. 879-89, 1999.

[11] H. Law , S. T. Neil, G. Dunn and A. P. Morrison, “Psychometric properties of the Questionnaire about the Process of Recovery (QPR),” *Schizophrenia Research,* vol. 156, no. 2-3, pp. 184-9, 2014.

[12] C. D. Taylor, P. E. Bee, J. Kelly and G. Haddock, “iMAgery-Focused Psychological Therapy for Persecutory Delusions in PSychosis (iMAPS): A Novel Treatment Approach,” *Cognitive and Behavioral Practice,* vol. 26, no. 3, pp. 575-588, 2019.

[13] D. Fowler, D. Freeman, B. Smith, E. Kuipers, P. Bebbington, H. Bashforth, S. Coker, J. Hodgekins, A. Gracie, G. Dunn and P. Garety, “The Brief Core Schema Scales (BCSS): psychometric properties and associations with paranoia and grandiosity in non-clinical and psychosis samples,” *Psychological Medicine,,* vol. 36, no. 6, pp. 749-759, 2003.

[14] M. Cloitre, M. Shevlin, C. R. Brewin, J. I. Bisson, N. P. Roberts, A. Maercker, T. Karatzias and P. Hylan, “The International Trauma Questionnaire: development of a self-report measure of ICD-11 PTSD and complex PTSD,” *Acta Psychiatrica Scandinavica,* vol. 138, no. 6, pp. 536-546, 2018.

[15] M. J. Power, “The structure of emotion: An empirical comparison of six models,” *Cognition and Emotion,* vol. 20, no. 5, p. 694–713, 2006.

[16] A. T. Beck, N. Epstein, G. Brown and R. A. Steer, “An inventory for measuring clinical anxiety: psychometric properties,” *Consulting and Clinical Psychology,* vol. 56, no. 6, pp. 893-7, 1988.

[17] D. Addington, J. Addington and E. Maticka-tyndale, “Assessing depression in schizophrenia: the Calgary Depression Scale,” *The British Journal of Psychiatry,* vol. 163, no. S22, pp. 39-44, 1993.

[18] R. Tennant, L. Hiller, R. Fishwick, S. Joseph, S. Weich, J. Parkinson, J. Secker and S. Stewart-Brown, “The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation,” *Health Qual Life Outcomes,* vol. 5, no. 1, p. 63, 2007.

[19] P. L. Morosini, L. Magliano, L. Brambilla and S. Ugoli, “Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning,” *Acta Psychiatrica Scandinavica,* vol. 101, no. 4, p. 323–329., 2001.

[20] J. Busner and S. D. Targum, “The clinical global impressions scale: applying a research tool in clinical practice,” *Psychiatry (edgmont),* vol. 4, no. 7, p. 28–37, 2007.

[21] I. R. White and S. G. Thompson, “Adjusting for partially missing baseline measurements in randomized trials,” *Statistics in Medicine,* vol. 24, p. 993–1007, 2004.

[22] “The manual Process and methods,” NICE health technology evaluation (PMG36), 2022.

[23] B. Janssen, A. Szende and J. Cabases, “Population Norms for the EQ-5D,” in *Self-Reported Population Health: An International Perspective based on EQ-5D*, Dordrecht (NL), Springer, 2014, p. Chapter 3.

[24] A. D. Keetharuth, D. Rowen, J. B. Bjorner and J. Brazier, “Estimating a Preference-Based Index for Mental Health From the Recovering Quality of Life Measure: Valuation of Recovering Quality of Life Utility Index,” *Value Health,* vol. 24, no. 2, pp. 281-290, 2021.

[25] S. R. Kay, A. Fiszbein and L. A. Opler, “The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia,” *Schizophrenia Bulletin,* vol. 13, no. 2, p. 261–276, 1987.

[26] S. Carr, A. Hardy and M. Fornells-Ambrojo, “The Trauma and Life Events (TALE) checklist: development of a tool for improving routine screening in people with psychosis,” *European Journal of Psychotraumatology,* vol. 9, no. 1, p. Article: 1512265, 2018.

[27] J. Williams, M. Leamy, F. Pesola, V. Bird, C. Le Boutillier and M. Slade, “Psychometric evaluation of the Questionnaire about the Process of Recovery (QPR),” *The British Journal of Psychiatry,* vol. 207, no. 6, p. 551–555, 2015.

[28] Rothwell, J C, S. A. Julious and C. L. Cooper, “A study of target effect sizes in randomised controlled trials published in the Health Technology Assessment Journal,” *Trials,* vol. 19, no. 1, p. 544, 2018.

[29] Pyle, M., Norrie, J., Schwannauer, M., Kingdon, D., Gumley, A., Turkington, D., Byrne, R., Syrett, S., MacLennan, G., Dudley, R., MacLeod, H. J., Griffiths, H., Bowe, S., Barnes, T. R. E., French, P., Hutton, P., Davies, L., and Morrison, A. P. (2016). "Design and protocol for the Focusing on Clozapine Unresponsive Symptoms (FOCUS) trial: a randomised controlled trial", *BMC Psychiatry*, vol. 16, 280, p1-12.

9 APPENDIX

Figure 1: CONSORT Diagram