Approved: 29.11.2023

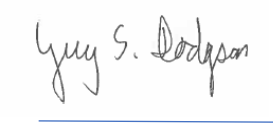
FULL TITLE	MUSE ARMS Feasibility RCT: A feasibility trial employing a prospective randomised, open-label, observer blinded, endpoint design assessing a targeted, computer/web based guided self-help psycho-education toolkit for distressing hallucinations (MUSE) + treatment as usual (TAU) compared to a time-matched TAU, offered by a multi-disciplinary team which includes emotional support, psychoeducation and stress management, aiming to reduce distress from hallucinations and improve functioning, in people with an At Risk Mental State (ARMS) for psychosis in UK secondary care mental health services.
REC number	23/NE/0032
ISRCTN number	ISRCTN58558617
IRAS number	323903
SAP VERSION	1.0
SAP VERSION DATE	29/11/2023
TRIAL STATISTICIAN	Akansha Singh
SENIOR STATISTICIAN	Jochen Einbeck
TRIAL CHIEF INVESTIGATOR	Guy Dodgson
TRIAL MANAGER AND DATA MANAGER	Jahnese Hamilton
SAP AUTHOR	Akansha Singh, Jahnese Hamilton, Jochen Einbeck

1. SAP Signatures

I give my approval for the attached SAP with Short Title “MUSE-ARMS Feasibility Trial” dated 29.11.2023

Chief Investigator

Dr Guy Dodgson

Signature: 

Date: 29/11/23

Statistician

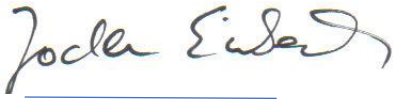
Dr Akansha Singh

Signature: 

Date: 29/11/2023

Senior Statistician

Prof Jochen Einbeck

Signature: 

Date: 29/11/2023

2. Table of Contents

Contents

3. Abbreviations and Definitions

AE	Adverse event
ARMS	At Risk Mental State for psychosis
CACE	Complier average causal effect
CBT	Cognitive behaviour therapy
CI	Chief Investigator
CNTW	Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust
CONSORT	Consolidated standards of reporting trials
CRF	Case report form
CSRI	Client Service Receipt Inventory (mental health)
EIP	Early Intervention in Psychosis
FEP	First Episode Psychosis
GEE	Generalised estimating equations
ITT	Intention-to-treat
LEAP	Lived experience advisory panel
ISRCTN	International Standard Randomised Controlled Trial Number
MHSDS	Mental Health Services Data Set
MUSE	Managing unusual sensory experiences
NICE	National Institute for Health & Care Excellence
NHS	National Health Service
PI	Principal Investigator

PIS	Participant Information sheet
REC	Research Ethics Committee
RCT	Randomised control trial
SAE	Serious adverse event
SAP	Statistical analysis plan
SAT	Statistical analysis team
TAU	Treatment as usual
TMG	Trial Management Group
TSC	Trial Steering Committee
TEWV	Tees Esk and Wear Valley NHS Foundation Trust

(List excludes abbreviations for measures and questionnaires as detailed in Sec 8.2-8.7).

4. Introduction

4.1 Preface

Psychotic disorders (such as schizophrenia) impose a huge public health burden, with hallucinations (hearing or seeing things that others do not) a significant and often highly distressing aspect of these disorders. Given the human costs of psychosis and the desirability of preventing its onset, therapeutic efforts have targeted groups considered at high risk, such as At Risk Mental State (ARMS). Our novel psychological intervention, Managing Unusual Sensory Experiences (MUSE), uniquely focuses on the varied experience of hallucinations. In an accessible, intuitive digital format, MUSE provides information to patients about the psychological mechanisms underlying their experiences, along with coping techniques targeting these factors to reduce distress. Already proving its value in first-episode psychosis (FEP) (Dudley et al., 2022), our approach may be particularly impactful in treating unusual experiences in ARMS (when interpretations of

unusual experiences are still highly changeable) thus reducing distress and potentially preventing transition to psychosis or other mental health conditions (Moritz et al., 2019). This work addresses Goal 1 of the recent shared goals for mental health research (Wykes et al., 2021) (namely to halve persistent mental health problems in children and young people). It could result in substantial benefits for patients and their families, the generation of new knowledge linking mechanism to hallucination phenomenology, reduction in pressure on services, and lessening of the societal cost of psychosis.

MUSE is unique in drawing on current, empirically supported psychological understandings to explain unusual experiences such as hallucinations. It exemplifies a novel approach to psychological therapy involving shorter, targeted treatments that use CBT-informed techniques to focus on specific symptoms. This has produced promising results for other psychosis symptoms such as paranoia (Freeman et al., 2016). MUSE is fully compatible with other important treatments such as CBT. This intervention also represents a significant technological innovation, potentially suited to a younger patient population who may be particularly open to material presented in a digital multimedia format, with potential for future adaptations for online administration. MUSE can easily be encompassed within current practice for future patient benefit.

A preliminary feasibility study (Dodgson et al., 2021) in ARMS showed promise in reducing symptoms and distress, including high acceptability and no adverse reactions. This study, however, lacked a control group, data on acceptability of randomisation, and outcome data to inform a sample-size calculation. These previous studies have allowed us to improve MUSE by incorporating a wider range of information about hallucinations (including visions), more co-produced materials, and development of specialist training.

Since MUSE works by targeting mechanisms underlying specific hallucination subtypes, it is of interest to investigate change in these mechanisms and their contribution to therapeutic efficacy. For example, hallucinations have been linked to internal experiences being erroneously attributed to events in the outside world (reality-monitoring) (Fernyhough et al., 2019). MUSE may help recipients to identify these confusions about the origins of their experiences, and the contributing roles of factors such as vividness of everyday inner experience, sleep and arousal (Dodgson et al., 2021).

4.2. Purpose of the analyses

This study will gather essential feasibility data on the deliverability of a fully powered trial in the future to reduce distress associated with unusual sensory experiences and improve global functioning in people with At Risk Mental State for psychosis through the delivery of brief targeted interventions. This research will also investigate the mechanisms assumed to be behind unusual sensory experiences, and the impact of the intervention, creating new scientific knowledge which can be used to refine the intervention and stratify treatment.

4.3. Checklist

The Appendix contains a checklist for this SAP against recommendations from the CONSORT 2010 guidelines (Eldridge et al., 2016).

5. Study objectives

The primary objective of this ISRCTN-registered feasibility randomised controlled trial is to resolve key feasibility uncertainties and inform the parameters of a future fully powered trial. This will include the collection of data concerning the feasibility of the recruitment target, as well as the preliminary effect of MUSE+TAU versus time matched TAU on general functioning and mental state in ARMS patients post therapy (12 weeks after randomization) and at follow-up eight weeks later (with some possible exceptions, see Section 6). Pooled standard deviations of the relevant outcome measures (which can be used in a sample size calculation for a future definitive trial) will also be calculated.

As secondary objectives, we aim to explore additional treatment effects on unusual sensory experiences, anxiety, depression, and quality of life, and whether there are indications of other factors (sleep disturbance and trauma) influencing treatment effects. Furthermore, we will test feasibility of collecting measures of psychological mechanisms, including psychological and personal (phenotypical) factors implicated in the clinical course of hallucinations, to inform a future investigation of whether any efficacy of MUSE is through impact on these mechanisms. Finally, we aim to collect routine data and

participant consent for a future records investigation testing feasibility of tracking transition to psychosis through medical databases (hospital records/Mental Health Services Data Set (MHSDS)), to examine which features of MUSE (presenting, treatment response and mechanistic) are most relevant to psychosis prevention.

6. General Study Design and Plan

This is a mixed-method feasibility trial employing a prospective, randomised, open-label, observer blinded endpoint design with MUSE+TAU compared to time matched TAU, with assessments at pre- and post-treatment (three-month follow up) and at five-month follow-up.

This is a multicentre trial taking place in NHS settings in the UK. The study will run through At-Risk Mental State services and Early Intervention in Psychosis services that provide an ARMS service. There are two NHS Investigator Sites with multiple services that serve populations across both rural and urban geographical locations in the North East and North Cumbria regions of England.

The study was initially devised for 9 months of recruitment. Since one of the two study sites was lagging behind in their recruitment, a possible contingency plan has been considered, three months into the trial, to add a 10th month of recruitment. If this was to happen, then due to time and funding constraints, the overall length of the study could not be extended. This means that, for individuals randomised in the last month of recruitment, the final follow-up would be at 16 weeks post-randomisation (that is, one month after the post-treatment measures corresponding to the primary outcome point were taken). In this circumstance, it is possible to account for any resulting time variation in the analysis; see Section 10.4.

Long-term outcomes (3 years post baseline) will be collected by the CI-led central research team via the Mental Health Services Data Set (MHSDS) or medical records if the MHSDS is unavailable. This long-term follow-up activity takes place after research sites have closed and is limited to data collection via the MHSDS/medical records, meaning participants have no direct involvement with researchers at this stage. Sites will submit NHS numbers of consenting participants to the CI or delegate for this follow-up

analysis. The CI or delegate will access personal data pertaining to the period from informed consent to 3 years after baseline assessment. The Assessing Transition to Psychosis study-specific CRF will measure transition to psychosis. The relevant data will be recorded and stored in a password-protected computer file for the follow-up analysis under a participant code. No personal identifiable information will be recorded. This analysis is considered external to this study and will therefore not be given further attention in this SAP.

6.1. Study intervention

The MUSE intervention (as amended) is a novel targeted, computer/web based guided self-help psycho-education toolkit and psychological treatment manual for managing distressing hallucinations in mental health, developed and owned jointly by Durham University and CNTW. Patients work with experienced therapists, under expert supervision, who utilise the MUSE package within therapy sessions to develop a formulation explaining the development of hallucinations and foster new skills and strategies for their management.

The treatment is divided into eight modules “What are Voices” / “How the mind works” / “Assessment” / “Inner Speech” / “Memory-based” / “Hypervigilance” / “Seeing Visions” and “Sleep”. Details about the specific content and purpose of each of these eight modules can be found in the Protocol.

6.2. Schedule of intervention

MUSE involves several weekly face-to-face sessions (~60min) of 6 core sessions with an option of two additional sessions. The number of sessions is based on previous work with other groups and feedback from ARMS therapists who currently employ MUSE in practice, but the clinician can choose to use the manual for more sessions, if they deem necessary. This design will ensure that participants receive adequate exposure to the manual in therapy sessions for us to determine its acceptability.

6.3 Comparison intervention: Treatment as usual (TAU)

To control for risk of bias from an undefined comparative treatment, and potential bias from dose effects, a time-matched TAU is included (Bosnjak Kuharic et al., 2019; Higgins & Green, 2011). In order to match the comparative brief intervention to practice within ARMS services, components of care were identified in an engagement meeting with ARMS service leads. These common core components, which could be described as Supportive Psychotherapy (needs based emotional support, psychoeducation, normalisation and stress management), were outlined as the interventions used by practitioners as part of their normal clinical toolkit, alongside routine multi-disciplinary care. We will investigate how frequently and consistently these supportive psychotherapy interventions are offered to inform whether these interventions could act as a comparator intervention in future trials. CBT is a core intervention, recommended by NICE guidance and offered across all services. CBT will form part of the care in both conditions. The number of sessions received by participants will be measured to investigate whether MUSE impacts on the number of sessions required (provided that these are reliably measurable, see also Sec 8.9).

Participation in this trial does not lead to the withholding of any treatment based on clinical judgement and we will record the interventions received within TAU in both conditions.

6.4. Participant timeline

Clinical teams supported by their Trust's clinical research delivery team members will identify potential participants from caseloads, clinics and newly accepted referrals, who will have a Perceptual Abnormality score (in CAARMS) of 3 or above in the last 4 weeks. Patients who potentially meet the eligibility criteria for the trial, and their parent/guardian where appropriate if under 18 years, will be informed of the study by a member of their clinical team (or a Trust clinical research staff member who works into the clinical team). Patients will be asked for their verbal consent to be contacted by a member of the research team and/or to receive further information on the study.

Prior to the informed consent meeting, the Participant Information Sheet (PIS) will be provided, and participants will be checked for eligibility via discussion with referring teams and in the participant-researcher discussion. The informed consent meeting will be

scheduled at least three days after the potential participant has received the PIS. Informed consent will involve review and discussion of the participant information with an authorised member of the research team who is delegated this duty by the Principal Investigator, or is the Principal Investigator at site. Participants will be informed that participation is voluntary and they can withdraw at any time without giving reason and without their medical care or legal rights being affected. There is no further post-consent eligibility testing.

Baseline measurements are taken following informed consent.

Assessment measures will be collected at baseline, post intervention (12 weeks post randomisation) and follow-up (20 weeks post randomisation), until all participants complete the follow-up assessment or withdraw.

Randomisation occurs after baseline assessments are completed. The 6-8 therapy sessions (MUSE / TAU) should be completed before the 12 week assessment point; we will record details of any participants whose therapy sessions run beyond the 12 week assessment.

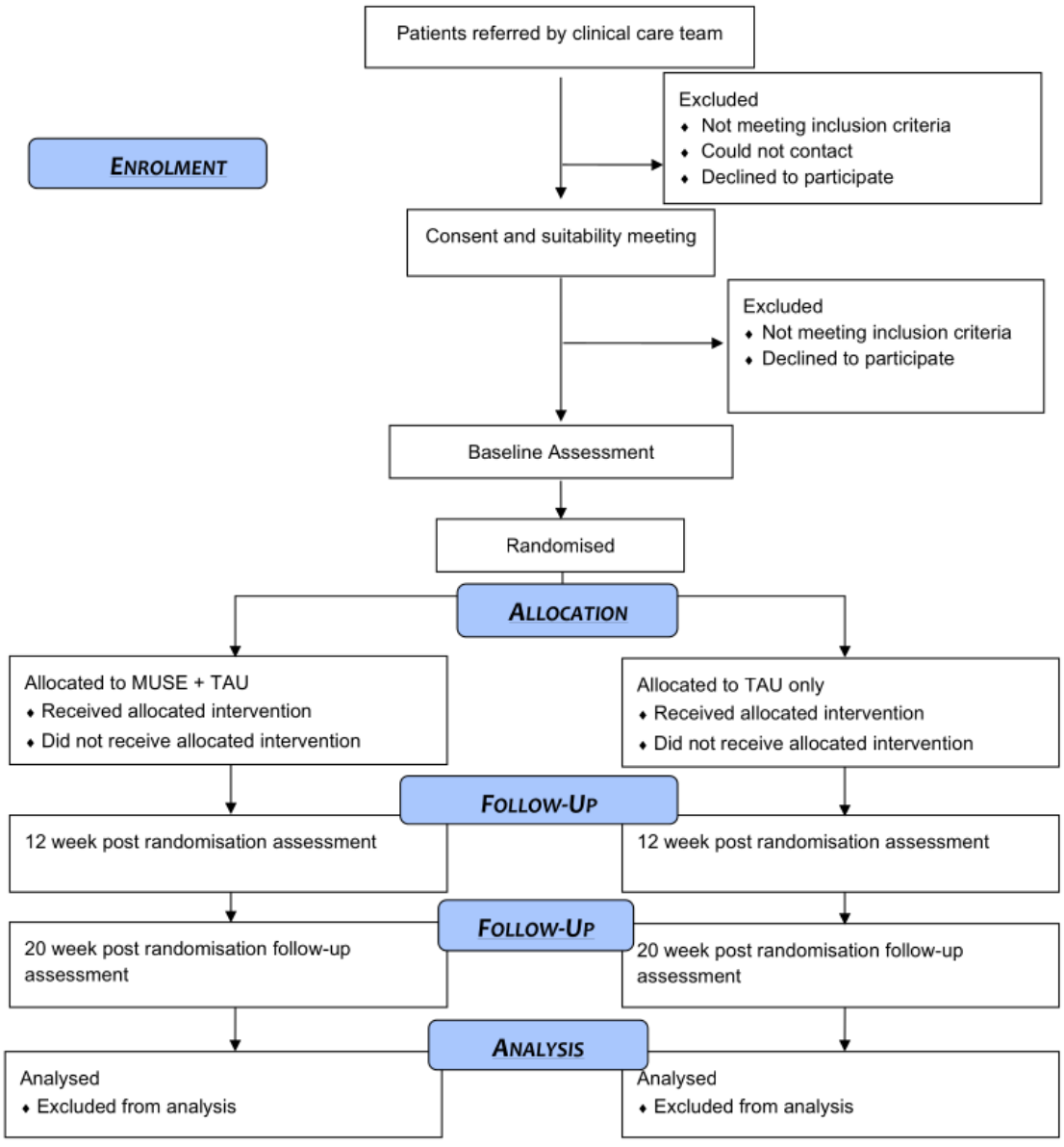
A participant is considered as withdrawn following a formally documented withdrawal actioned at the request of the participant. This has four levels (Withdrawal from Intervention/Withdrawal from trial assessments/ Withdrawal from long-term follow-up/ Withdrawal from selection for qualitative interview) and reasons are reported in Section 4 of the Withdrawal Form.

A participant is considered a drop-out where the participant has not attended their follow up appointment (at week 20, or week 12 and 20), and the participant has not indicated their intention or reason for disengaging.

Participants who are discharged from an ARMS service/pathway will remain in the trial for trial assessments. Data on whether participants have been discharged and reason for discharge will be collected at 20 weeks.

6.5. Study Flow Charts

Figure 1. Trial flow diagram (IRAS ID: 323903/ Version: 2.0/ Date:24.02.2023)



6.6. Inclusion Criteria

- in contact with an ARMS service or accepted on an ARMS pathway by EIP services
- aged 14–35
- hallucinations / unusual sensory experiences scoring at least 3 on the Perceptual Abnormalities Subscale of the CAARMS
- hallucinations considered by the patient to be a key target problem
- judged to have been clinically stable for the preceding 2 weeks

6.7. Exclusion Criteria

- intellectual disability or severe cognitive dysfunction affecting ability to engage with research materials
- lacking capacity to give informed consent

7. Randomisation and Blinding

7.1 Randomisation

Eligible participants who have completed baseline assessments will be randomised. An independent web-based randomisation service (sealedenvelope.com) is used for the trial. Randomisation will be in the ratio 1:1 to the two groups: MUSE+TAU (intervention) or TAU (control). Randomisation will be stratified by site, gender (M/F/Other) and age (14–17 years/18–35 years inclusive). Randomisation allocation will be independent and dynamically generated using a randomised modified minimisation method (Kuznetsova & Tymofyeyev, 2012) to assure allocation concealment along with preservation of allocation ratio.

The researcher who completes the baseline assessment will enroll participants for randomisation, and Sealed Envelope will assign participants to the two groups. Randomisation allocation is made known to the CI and site PIs, the Trial Coordinator(s) and the trial therapists only at the point of randomisation, by email. Research assessors for the trial will be blind to the allocation throughout the trial. The unblinded trial coordinator or local site lead delegated to this role will inform participants of which group they been allocated to.

7.2 Blinding

Clinicians, therapists and participants will be unblind. Research outcome assessors (for the trial measures and cognitive tasks) will be blind. Maintaining blindness of research assessors is crucial, and care will be taken within the research team to avoid accidentally unblinding outcome assessors. Any cases of inadvertent unblinding will be discussed in a TMG, and TSC will monitor unblindings by each site regularly and implement corrective action if necessary. Participants and clinical teams will be reminded prior to each assessment timepoint by the research team that they must not inform the blinded researchers of their group allocation. Where unblinding occurs during the assessment, no further assessments will be taken and another appointment will be made for a blinded member of the research team to complete the measures.

Trial statisticians will be partially blind. This means they will be blind for all primary analyses and almost all secondary analysis, except those analyses which involve data relating directly to the respective interventions, such as the number of MUSE sessions (as clearly this requires knowledge that these participants were part of the MUSE intervention arm). In practice, this will be handled by providing blinded data to the Statisticians initially. Once all analyses with the blinded data (including the analysis of the follow-up data) have been carried out and confirmed by the TMG, the unblinded information will be handed to the Statisticians.

After randomisation, the Client Service Receipt Inventory (CSRI) (mental health) service use data (incorporating use of MUSE in either treatment arm), Transition to Psychosis data, and Adverse Event data will be collected from medical records by an unblinded member of the research team. This is because it is not possible to access medical records data without becoming unblinded to the treatment allocation.

The Trial Coordinator will oversee the maintenance of blinding and will monitor any blinding breaches closely. Any unblinding of blinded assessors will be classed as a protocol deviation and reviewed by the site PI and Trial CI. The TSC will monitor any unblinding and implement corrective action if necessary.

8. Study Variables

8.1. Demographic variables

Raw demographic variables recorded from participants are as follows:

- Age at baseline (quasi-continuous, in years)
- Self-reported Gender (Male/Female/Other)
- Ethnicity
- Living situation (living with parents / living alone / living with partner / living with others)
- NEET (Not in Education, Employment, or Training) status: Y/N
- Other socioeconomic indicators
- Service use: Participants use of services is collected at baseline, 12 weeks, and 20 weeks using CSRI, as amended for the trial. This will collect data on primary care, community care, hospital, and criminal justice service use. Medication use is also collected for medications for mental health. See Section 8.2 for additional trial-related information collected from CSRI.

8.2 Stratification and trial-related variables

In the list which follows, variables 1.-3. are used as stratification variables, and also as covariates in all models in the quantitative analysis. Variable 4 will be required for analysis of the follow-up outcome measurements.

1. A binary variable (<18ys, >18ys) created from the age variable.
2. A variable (Male/Female/Other) for gender [3 levels].
3. A categorical variable for site [2 levels].
4. A time variable indicating when the measurement was taken: This variable can take at most four distinct values: 0 weeks (baseline), 12 weeks (post-test), 20 weeks (follow-up), as well as possibly 16 weeks, should the trial require an extension to the recruitment time window, with some participants having their

follow-up assessment early. No other values of this variable will be permitted, as doing so for individual patients may lead to accidental unblinding. Values will be assigned according to the intention of taking the measurements at a certain time point, even if that time point cannot be held for logistic or other matters.

5. The CSRI will contain trial-specific questions capturing use of CBT, and use of MUSE, regardless of allocation group (also use of group therapy, other psychological therapy).
6. Further information will be available from therapy packs, such as on the type of TAU received (CBT session Y/N, CBT assessment, Formulation, Needs-based emotional support, Social Support, Normalisation, Stress management, Psychoeducation), or reasons for discharge.
7. Other relevant trial information, such as Adverse Events (AEs), will also be extracted from the participants' case report form (CRF).

8.3. Primary outcome measures: Trial feasibility outcomes

The primary outcome of the trial, which also directly links to the progression criteria (Section 11.2), and therein referred to as "Recruitment", is the "number of participants consented and randomised", per month, over all sites. We will report these monthly recruitment rates jointly for all sites, and separately by site. We will in all cases also report the corresponding total numbers (not averaged by month).

In order to inform the progression criteria (Section 11.2), and with view of the CONSORT graph (Figure 1), we will also record the

- number of MUSE sessions completed by each participant randomised to MUSE (participants who complete at least 4 out of 8 sessions will be labelled as "(treatment) compliant", and we are interested in the proportion of compliant participants to inform therapy engagement);
- whether each participant completed the primary outcome measure at primary assessment endpoint (we are interested in the proportion of such completions to inform assessment retention);
- whether each participant rated the therapy as acceptable or not (the proportion of acceptable ratings will inform therapy fidelity);
- the number of MUSE-related SAEs (this will inform the Safety measure)

For reference, the recruitment and attrition pipeline anticipated in the funding bid is as follows:

Table 1: Recruitment and attrition according to Funding bid.

Sites	ARMS cases pa*	ARMS cases pcm**	Meeting criteria (assuming 65%)	Estimated recruitment pcm** (assuming 50%)	Estimated completion pcm** (assuming 80%)
TEWV	173	14.4	9.4	4.7	3.8
CNTW	189	15.8	10.3	5.1	4.1
Total	329	27.4	19.7	9.8	7.9

The progression criteria (towards a definitive trial) provided in Section 11.2 of this SAP relate directly to these numbers.

8.4 Primary outcome measures: Treatment effects

The primary outcome measures, to be completed at each time point (baseline, post-intervention, follow-up) are

- **SOFAS** (Social and Occupational Functional Assessment Scale; Goldman et al., 1992). The SOFAS is a clinician/clinical researcher rated single-item scale to assess social and occupational functioning. The period of assessment for this trial is the last two weeks. Scoring is from 0 to 100; higher scores represent better functioning. Scoring is completed independently of patient-reported psychological symptoms; however, impairment due to ill health (physical and mental) is scored, whereas impairment due to environmental factors or lack of opportunity is not scored.
- **PSYRATS** (Psychotic Symptom Rating Scales; Haddock et al., 1999). The PSYRATS is a clinician/ clinical researcher administered semi-structured interview of hallucinations (11 items) and delusions (six items). Each item is rated by the interviewer on a 5-point nominal scale (0–4). The PSYRATS auditory hallucinations total score will be used to assess the severity of hallucinations. Sums of scores across the hallucinations subscales (Woodward et al., 2014) of Distress (questions 6,7,8,9 &11) and Attribution (questions 3 & 5) will be analysed with equal importance as key problem areas which MUSE is seeking to target.

Table 2. Main Outcome Measures

Outcomes	Focus	Measure
Primary Global State Outcome	Functioning	Social and Occupational Functional Assessment Scale (Goldman et al., 1992)
Primary Mental State Outcome	Mental State	Psychotic Symptom Rating Scale hallucinations total (Haddock et al., 1999)
	Hallucinations	Psychotic Symptom Rating Scale distress (Haddock et al., 1999; Woodward et al., 2014)
	Attribution	Psychotic Symptom Rating Scale attribution (Haddock et al., 1999; Woodward et al., 2014)

8.5 Secondary outcome measures: Treatment effects

- **PSYRATS Delusions.** The PSYRATS Delusions Rating Scale is a subscale of the PYSRATS (see Section 8.4) designed to gauge the intensity of psychotic symptoms, with a particular emphasis on delusions (beliefs), consisting of six items rated by the interviewer on a 5-point nominal scale (0–4). The total score across the six items will be used.
- **CAARMS** (Comprehensive Assessment of At-Risk Mental States; Yung et al., 2005). The CAARMS is a clinician/clinical researcher administered semi-structured interview commonly used to assess patients referred to ARMS services. This trial uses the CAARMS-PA subscale (Perceptual Abnormalities) part of this assessment. The PA scores encompass a single item score for each of the following; (i) Global Rating Scale, rated between 0-6. (ii) Frequency and duration, rated between 0-6. (iii) Pattern of symptoms in relation to substance misuse, rated 0-3, and (iv) Level of Distress (In Relation to Symptoms) rated between 0-100. This

subscale data can be obtained from participants' interviews or from medical notes. The analysis will be carried out on each subscale separately.

- **GAD-7** (General Anxiety Disorder, Spitzer et al., 2006) is a brief self-report questionnaire of 7 items, and a categorical impact on functioning question (not difficult at all, somewhat difficult, very difficult, extremely difficult), used as a screening tool for anxiety, which has good reliability and validity. The total scores across all the seven items will be used.
- **PHQ-9** (Patient Health Questionnaire, Kroenke et al., 2001) is a brief, self-report questionnaire of 9 items, and a categorical impact on functioning question (not difficult at all, somewhat difficult, very difficult, extremely difficult), to measure depression symptom severity with good reliability and validity (Kroenke et al., 2001).
- **MMHS** (Multi-Modal Hallucinations Scale; Dudley et al., in prep.) is used to assess cross-modal sensory experiences. The MMHS is a brief self-report measure which assesses unusual sensory experiences in six modalities: auditory, visual, olfactory, gustatory, bodily sensations and sensed presence. The measure asks questions about the frequency and distress caused by these experiences and asks for a brief description. The MMHS is yet unvalidated, but has been included as it investigates whether unusual sensory experiences are combined (such as seeing a vision which is the source of a voice).
- **ReQoL-20** (Recovering Quality of Life; Keetharuth et al., 2018) is a brief self-report 20 item questionnaire and will be used to measure quality of life. Questions are rated using a 5-point nominal scale (0-4) and a mixture of reverse scores items (scored correctly at point of data collection via Qualtrics), with higher scores reflecting better quality of life, and an increase of 5 points denotes reliable improvement in quality of life, whereas a decrease of 10 points denotes a deterioration in quality of life. Keetharuth et al. (2021) carried out an item response theory analysis of the measure and found that it has robust internal construct validity.

8.6 Secondary outcome measures: Moderators

Moderators are important baseline measures obtained prior to randomisation (Pincus et al., 2011) to assess whether the treatment is more or less effective from participants with certain characteristics such as higher levels of hallucinations, insomnia, or trauma symptoms.

- **CAARMS-**PA (measure as explained in the previous subsection) as a moderator will use the baseline sum score of the Global Rating Scale, combined with Frequency and Duration score (total score will be between 0 and 12). This will show a scale of severity of hallucinatory experience. As the population of the study is an 'at risk mental state' population, it is possible that the individuals with more severe symptoms tend to benefit more from MUSE (for instance, as they are more likely to complete all sessions; Dodgson et al., 2021b). A moderation analysis will help indicate if this is the case, or not.
- **ISI** (Insomnia Severity Index; Bastien et al., 2001) is a brief self-report questionnaire of 7 items to assess sleep difficulties and severity of insomnia. Users answer questions relating to their quality of sleep and levels of insomnia over the past 2 weeks, which are rated using a 5-point nominal scale (0–4). Scores of 15 or above are indicative of clinical levels of insomnia, with scores between 8 and 14 being indicative of subthreshold insomnia. The ISI has excellent internal consistency (Cronbach's alpha= .92) (Gagnon et al., 2013), and has been successfully implemented to assess insomnia in patients with psychotic disorders in previous research (Miller et al., 2019). The total score for the ISI at baseline will be used. Insomnia has been demonstrated to show causal effect on hallucinatory experiences (Freeman et al., 2017). Whether the MUSE intervention, which contains one module on sleep (Dodgson et al., 2021b), or TAU in ARMS services are providing adequate intervention for individuals with high levels of sleep disturbance (and unusual sensory experiences) is unknown and requires further investigation.
- **Trauma:** Two trauma self-report scales are used in the trial, participants complete one scale depending on participants age (**ITQ** (International Trauma Questionnaire); Cloitre et al., 2018/ **ITQ-CA** (International Trauma Questionnaire - Child and Adolescent Version); Cloitre et al., 2018; Haselgruber et al., 2020)

completed for participants aged 14-17 years at baseline). Due to differences in question style between the two questionnaires participant scores will be scaled to allow combined analysis. Here the raw total score for each participant will be scaled by dividing the score by the total possible score and multiplying by 10. Whether the MUSE intervention, which contains one module on traumatic memories and intrusions in relation to unusual experiences (Dodgson et al. 2021b), or TAU in ARMS services are providing adequate intervention for individuals with high levels of trauma (with unusual sensory experiences) is unknown and requires further investigation.

- **Preferences for psychological therapy or support** (non-validated study-specific questionnaire) is obtained at baseline only. The questionnaire contains 13 questions addressing four domains of preference, (i) the number of sessions, (4 range options and don't know) (ii) the focus of the therapy (7 questions, ranked not important, somewhat important, very important for; includes medication, includes talking therapy, addresses feelings of anxiety, feelings of low mood, helps understand usual sensory experiences (USE), helps manage USE, helps feel less distressed by USE), (iii) therapist style of engagement (4 questions, ranked not important, somewhat important, very important for; being given space to talk and feel heard, work with therapist to make sense of experiences, involved setting own goals, given new ideas of how to cope with experiences), and (iv) preference for MUSE treatment (Y/N). Preferences for psychological therapies have shown strength of relationship with outcomes in some settings (e.g., Cooper et al., 2018; Williams et al., 2016).

8.7 Secondary outcome measures: Mechanisms

We investigate which psychological mechanisms (Inner speech, Memory, Hypervigilance, Visions) are influenced by the treatment and contribute to its clinical effect, thus informing a future investigation of whether any efficacy of MUSE is through impact of these mechanisms. Assessments related to measure include a self-report questionnaire and accompanying computerised cognitive psychology tasks. Specifically, the mechanisms and associated measures used are

- For **Inner Speech**, the Varieties of Inner Speech Questionnaire-revised (Alderson-Day et al., 2018), scores for each of the five factors (dialogic, evaluative/critical, other people, condensed and positive/regulatory, and the Auditory signal detection task (Moseley et al., 2021);
- For **Memory**, the Dissociative Experiences Scale-Brief (DES-B; Dalenberg and Carlson, 2010), average score, and the Inhibition of Currently Irrelevant Memories task (Paulik et al., 2007);
- For **Hypervigilance**, the State-Trait Anxiety Inventory-Short Form (Spielberger, 1983) average score for state, and average score for trait, and the Jumbled Speech Task (Fernyhough et al., 2007);
- For **Visual**, the Plymouth sensory imagery Questionnaire-SF (Andrade et al., 2014) visual subsection (5 questions) average score, the Visual signal detection and the Face pareidolia Task (Smailes et al., 2020).

Participants will only complete the above subtype-specific measures for a maximum of two subtypes.

8.8 Other secondary outcome measures

Other secondary outcomes include:

- **Transition to psychosis**: This will be confirmed via CI review of discrete criteria being met (based on the criteria used in the IPPACT study), however in the short follow-up period of 20 weeks it is not expected that this number will be high. The long-term follow-up (3 years) from the Mental Health Services Data set will investigate this further.
- **Adverse events**: Adverse events relating to psychological wellbeing will be recorded and reported for the novel intervention and comparison treatment arms of the trial.
- **Serious adverse events**: Serious adverse events (and their relatedness and expectedness) will be recorded and reported for the novel intervention and comparison treatment arms of the trial.
- **Urgent safety measures**: Any urgent safety measures, if occurred or none occurred, will be reported.

- **Impact of MUSE on TAU:** A further outcome will be whether offering MUSE impacts on usual care: whether it reduces the length of CBT interventions needed, whether there are differences in rates of discharge or reasons for discharge; and other treatment use. Details of all treatment received in both groups will be recorded using CSRI questions 4 to 5, amended for this trial to add specific questions to measure receipt of relevant interventions in both arms for the duration of the study. See also Section 8.9 concerning the feasibility of collecting data on attended CBT sessions.
- **Treatment integrity** (adherence) will be assessed by sessions checklist data and independently (10% sample) of fidelity checks of audio recordings of sessions.
- The **therapeutic alliance** and responsiveness of participants as reflected on by therapists and participants will be assessed qualitatively by STTS-R data (Revised version of the Satisfaction with Therapy and Therapist Scale; post-intervention only).

8.9 Additional feasibility questions

This subsection bundles items which are not considered outcomes in their own right, but which still provide useful information to assess aspects of the feasibility of the delivery of the trial, and the quality of the data on the outcomes presented in previous subsections, with particular view to the delivery of future trials.

For most questions listed here, no formal statistical analysis (apart from simple descriptive where applicable) will be undertaken; the focus is on establishing whether the respective data/information could be obtained or not.

Feasibility of defining TAU. We will assess the core components delivered in TAU using data from the TAU therapist packs and CSRI data.

For individuals randomised to TAU,

1. Completeness of TAU Therapist packs
2. Availability of data for participants where Therapist packs are not available

For individuals randomised to either MUSE of Control,

- What data on additional TAU components e.g. Psychiatrist visits, Community Psychiatric Nurse input, are captured on the CSRI and how do the allocated groups differ in regards to this?

Feasibility of measuring the number of CBT sessions. We will assess the feasibility of collecting the following information.

For a given individual randomised to either MUSE or Control,

1. Did the individual receive CBT?
2. If so, how many sessions of CBT?
3. At the final assessment/data collection from medical notes, was CBT considered completed (yes/no/unknown)?
4. Hence, will it be meaningful to say, "the individual completed x CBT sessions", enabling us to examine the impact of MUSE on the number of CBT sessions offered in this group (see Sec 10.8)?

Feasibility of measuring contamination of TAU with MUSE. For a given individual randomised to Control,

- Can we measure the degree of contamination of TAU with MUSE, if any?

Feasibility of collecting mechanisms data. With view to the inclusion of mechanism tasks in a future trial, we will report on the following:

- The number of mechanism sessions, for each subtype, across all patients (by treatment arm, site, and time point);
- Our ability to collect data from each of the four types of mechanisms at each time point, specifically.
    - Were there any limitations on the ability of sites to deliver the collection of mechanism data?
    - Were there limitations in the acceptability of the mechanism assessments to patients (such as patients unwilling or uninterested to engage with a certain mechanism task; especially when doing it "again" at follow-up);
- Related questionnaire completion rates/ missing data;
- Feedback from researcher assessors on the feasibility of taking these measurements;

- Quality of data obtained from the cognitive tasks, including hits and false alarms (checking also for floor and ceiling effects to consider callibration requirements) as outcome variables for the auditory and visual signal detection tasks;
- Means and standard deviations on the key mechanisms outcome data;
- Variability in these data across time points;
- Any impact of changing task administration method on data completeness;
- Any technological problems that arose;
- Establishing criteria for including particular mechanisms measures in future trials.

8.10 Schedule of measurements

Table 3: Trial Assessments and Key Participant Procedures Schedule, with timing of study measurements.

Assessments/ procedures	**Participant identification**	**Enrolment & baseline**	**Randomisation**	**Intervention Weeks 1-12**	**12 weeks post randomisation (+/-10days)**	**20 weeks post randomisation (+/-10days)**
Recruitment and eligibility discussions	X					
Informed consent		X				
CSRI Sociodemographic Q1-3.5		X				
Randomisation			X			
MUSE & TAU / TAU Intervention				◄——►		
Blinded assessments						
MUSE ARMS Primary Outcome Measures: SOFAS & PSYRATS		X			X	X
CSRI service use Q4.1-4.4		X				
CSRI Q4.5 criminal justice services and Q5 medication		X			X	X
MUSE ARMS Secondary Outcome Measures: CAARMS-PA, PHQ-9, GAD-7, ReQoL-20, ISI, ITQ/ITQ-CA, MMHQ		X			X	X
Subtype Measures & Cognitive Tasks*[1] (1-2 subtypes selected per participant): See Table 3		X			X	X

Treatment preference		X				
Unblinded assessments						
CSRI service use at follow-up Q4.1-4.4					X	X
Transition to Psychosis data					X	X
Adverse Event (AE) data					X	X
Therapeutic Alliance STTS-R					X	
Participants interviews (Withdrawals sub-sample)				←	→	
Participants interviews (MUSE completers sub-sample)					← →	
Participants interviews (TAU sub-sample)					← →	
Therapists interviews (sub-sample)					← →	

9. Sample Size

A formal sample size calculation has not been performed for this feasibility study. Our goal is not to assess treatment effectiveness but to establish whether we can undertake future pilot and definitive studies to address effectiveness. We have a target of recruiting 88 individuals over the recruitment period. Based on past research of psychological therapy conducted in the Northeast, we have estimated attrition of 20% meaning approximately 70 people will complete the study. Guidance on external pilot studies indicates that samples of 35 per arm or more give a reliable estimate of the standard deviation of the outcome measure (Teare et al., 2014); however, the aim of the present study is not primarily to generate parameter estimates for a full trial, but to establish MUSE's acceptability and feasibility.

10. Statistical plans

Analyses will follow intention-to-treat (ITT) principles, with data analysed according to randomisation irrespective of treatment received. Cases with missing values/drop-outs will only be removed where this is necessary for the specific analysis to be carried out. Analysis will be carried out via the latest version of R, with models fitted using the glm, lmer, glmer, or gee functions as appropriate.

All analyses up to (and including) Section 10.7 will be blinded to treatment group identity. Primary endpoint analysis will occur after data lock of the trial for the 12 week assessments. Secondary endpoint analysis will occur after the data lock for the 20 week assessments. See Section 10.11 for further details on Timing of analysis and blinding of Statisticians.

10.1 Descriptive analyses

Recruitment numbers (total, per site, and per month) as well the various progression indicators required for use in Section 11.2, will be calculated according to the descriptions given in Section 8.3.

All other data from Section 8.1 to 8.7, i.e.,

- demographic variables (including medication use);
- stratification and trial-related variables;
- all primary and secondary outcome variables, including all moderators and mechanisms,

will be summarised, for control and intervention separately (wherever this is feasible), and at each of baseline, post-treatment, and follow-up, using mean±standard deviation and median±interquartile range for continuous data; frequency and percentages for binary or categorical data; and rates for count data.

Additionally, for all outcome measures (Sec 8.4 to 8.7), pooled standard deviations across treatment arms will be calculated. These standard deviations will be of relevance for the sample size calculations for a future trial.

The descriptive analysis will also identify whether there are differences in baseline characteristics between sites (such as age, gender, SOFAS and PSYRATS scores).

Hallucinations subtype selections will be summarized, including rates of agreement between researcher assessors and therapists on subtypes.

Preferences for psychological therapy or support data will be summarised.

We will pilot the ReQoL-Utility Index with the ReQoL data for health economic analysis calculation. The feasibility of using the ReQoL-Utility Index for descriptive summaries by treatment groups will be investigated.

Further descriptive analysis covering elements of items 5 (CSRI), 6 (therapy packs) and 7 (Case report forms) from Section 8.2., will be delegated to the unblinded analyses in Sections 10.8 and 10.9.

10.2 Difference-in-difference analysis

The analyses described in this subsection will be carried out for the following outcome variables:

- all primary outcome variables (Section 8.4);
- all secondary outcome variables (Section 8.5);
- the moderator variable ISI (Insomnia; Section 8.6);
- the mechanism variables (Section 8.7), where feasible.

We will calculate, at each of post-treatment and follow-up, the differences of the outcome measurement to the baseline measurements, yielding sets of differences of observed outcomes to baseline. We will visualize, for each outcome and measurement point separately, these sets of differences between treatment arms in boxplots, hence allowing a visual inspection of the signal of efficacy of the treatment.

We will also, for each outcome type and measurement point, and for each treatment arm, compute the mean value for each of these differences, as well as their difference between the treatment arms (this is a classical “difference-in-difference” estimator).

We will furthermore, for each outcome type and measurement point, compute the Welch t-statistic and (unadjusted) effect sizes between the treatment arms. While these quantities will allow a quantitative assessment of signal of efficacy, we will stop short of computing p-values or formally assessing significance. For the mechanisms, the ability to do carry out these tests is to some extent a feasibility question, since sample sizes for some of the mechanisms may be small. Furthermore, mechanisms which are insensitive to the treatment are unlikely to act as mediators, which is of relevance with view to the design of a future trial. Hence, it is important to carry out these analyses at this stage, while bearing in mind that the study is not powered or designed to prove significance of effects.

10.3 Models involving baseline and post-treatment (12week assessments) only

The analyses described in this subsection will be carried out for the following outcome variables:

- all primary outcome variables (Section 8.4);
- all secondary outcome variables (Section 8.5).

For these analyses, which do not involve the follow-up measurements, the baseline assessment will be treated as a covariate. The following additional covariates will be included: age, gender, and site (as a fixed effect, due to the small number of sites involved).

The models will contain a treatment indicator in order to identify the MUSE effect. The object of interest is the estimated coefficient corresponding to this indicator, as well as the standard errors of this estimate.

The effect of MUSE will be estimated using generalised linear models with the appropriate distribution and link function. Normal distributions with identity link will be used for continuous outcomes, logistic regression models for binary or categorical outcomes, and negative-binomial distributions with log link for count data outcomes.

That is, denote p_i the post-test measurement, b_i the baseline measurement, m_i a treatment indicator which takes the values of 1 and 0 for the two allocations while maintaining the blinding. Then the model takes the shape

$$g\big(E(p_i)\big) = \beta_0 + \beta_1 m_i + \beta_2 b_i + x_i^T \beta$$

where x_i is a vector of further covariates (coded factors for age, gender, and site), β_0, β_1, β_2, and β are coefficients, and g is is an appropriate link function (which is the identity for normal response, the logit for categorical or binary response, and the log for count data response). The object of interest will usually be the estimator $\hat{\beta}_1$ of β_1, which, together with its standard error $SE(\hat{\beta}_1)$, will allow the computation of t-values $\hat{\beta}_1 / SE(\hat{\beta}_1)$ and hence help assessing any signal of efficacy. Effect sizes $\hat{\beta}_1 / \sigma$, where σ is an estimate of the residual standard error obtained from the fitted model, will also be computed. Further inference on this effect size, such as establishing its significance, will not be attempted.

10.4 Models involving baseline, post-treatment, and follow-up

The analyses described in this subsection will be carried out for the following outcome variables:

- all primary outcome variables (Section 8.4);
- all secondary outcome variables (Section 8.5).

This is now a longitudinal data scenario. For each individual, we have up to three measurements at different time points. We calibrate these time points such that the baseline measurement is always taken at time=0, and time=12, 16, or 20 weeks depending on how many weeks after baseline the respective post-treatment or follow-up measurement are taken. This is a repeated measures layout, where within-subject correlations over time need to be accounted for through an additive random effect.

Denote the time index by $t = 0$ (baseline), $t = 1$ (post-treatment) and $t = 2$ (follow-up). Then a time variable is defined as

$time_{it}$= Number of weeks between measurement t and the baseline measure, for participant i

Clearly, this implies by construction $time_{i0} = 0$.

The models contain furthermore the same covariates as in Section 10.3, that is the factors for age, gender, site, and the treatment indicator, but no baseline assessment since this is now incorporated as an outcome at time index t=0.

Define the outcome variable at index t as y_{it}, $t = 0,1,2$. So, in the notation from Section 10.3, we would have $y_{i0} = b_i$ and $y_{i1} = p_i$.

The model now takes the shape

$$g\big(E(y_{it})\big) = \beta_0 + \beta_1 time_{it} + \beta_2 m_i \times 1_{\{t=1\}} + \beta_3 m_i \times 1_{\{t=2\}} + x_i^T \beta + u_i$$

with u_i denoting a subject-specific random effect. The objects of interest are now the estimates of β_2, β_3 and their standard errors. The previous comments on effect size apply accordingly.

This is a generalised linear mixed effect model, which will be estimated using the appropriate distribution and link functions using model estimation methodology under Gaussian random effects. For binary or categorical outcomes, the estimation will be carried out using generalised estimating equations (GEE: Dahmen and Ziegler, 2004). The mixed-effects models and GEE account for the repeated measurements per participant over the follow-up time points.

10.5 Missing data

If the post-treatment outcome is missing for any analysis in Section 10.3, then this case is dropped from the analysis as there is no possibility to adjust for this. The underlying assumption justifying this approach is that the rate of unit drop-outs is independent of treatment status. We will provide rates of missing data for the primary outcome variables, which will be calculated from the proportion of participants with available data at baseline, post-treatment and follow-up. Using cross-tabulation, we will assess whether there is evidence of non-random drop-out or missingness by comparing proportions of drop-outs and missing outcomes in each treatment group.

Notably the longitudinal analysis in Section 10.4 is robust to missingness of outcomes; that is if for a specific participant any of baseline, post-treatment, or follow-up are missing,

then the remaining observations from that case can still be used in the analysis and will contribute to the power of the fitted model. Hence, no adjustment for missingness (of outcomes) is required in this case.

If there are any instances of missing covariates, then the corresponding cases will initially be removed from the analysis. Subsequently they will be imputed (using Multiple Imputation techniques) and the robustness of the conclusions under the use of these imputations will be studied, as a sensitivity analysis.

10.6 Moderation analysis

For the primary outcomes only (Section 8.4), we will assess at both post-treatment and follow-up, the impact of the moderators listed in Section 8.6. Specifically, for the moderators CAARMS-PA, ISI, and Trauma, we will check whether these affect the impact of the treatment on the outcome. This will be achieved by adding the respective moderator as a predictor variable (to a model of the type as displayed in Section 10.4), along with an interaction term for treatment and the moderator. For the preferences questions, we will analyse the strength of relationship with the primary outcomes for each allocation group.

Additionally, we will carry out a secondary analysis in which age is included as a continuous predictor variable (see Section 8.2), and with an interaction term for treatment and age, in order to assess whether age affects the primary outcome variables in some systematic fashion, and whether this dependency is impacted in some way by the treatment.

10.7 Mediation analysis

In addition to estimating the difference between the MUSE+TAU and TAU groups, structural equation models will be used to estimate the average causal mediation effects (ACME) and to examine how the different mechanism components (see Section 8.7) mediate the estimated impact of MUSE on the primary outcomes (see Section 8.4). The mediation analysis may also inform candidate mechanisms and dose of treatment for

future trials. The R package mediation will be used for this analysis. The mediation analysis will be carried out for post-test and follow-up separately.

10.8. Adherence to treatment

The analysis in this subsection is carried out after unblinding of the Statisticians.

We initially report *treatment allocation adherence*, which means whether patients do engage at all with the treatment that they were randomised to, or whether they insist on (or somehow get access to) the other treatment. This is hence for each patient either TRUE or FALSE. Total numbers and proportions of treatment allocation adherence will be reported for both treatment arms.

For the MUSE arm only, we additionally consider *treatment adherence.* Treatment adherence measures the degree to which they engage with the randomised (MUSE) treatment, in terms of the number of sessions attended. Hence, treatment adherence is 0 if treatment allocation adherence is FALSE and is a value between 1 and 8 otherwise (it is noted that some participants may do more than 8 MUSE sessions, but these will still be considered as 8 for this analysis). Additionally, we consider the MUSE Treatment as *compliant* (also called: "treatment compliant") if four or more sessions of MUSE have been completed, so compliance is a function of treatment adherence. *Therapy engagement* is then the proportion of compliant participants. All this information will be presented descriptively. We will also present information on the modules that have been used by the therapists.

We will assess the effect of treatment adherence, the number of MUSE sessions on the primary outcomes SOFAS and PSYRATS (see Table 2) using a regression model, where the number of completed MUSE sessions will be modelled as a fixed effect covariate. Other covariates except site will be omitted for this analysis. Again, ITT will be followed in this analysis, which means that participants who were randomised to the MUSE treatment group but did not adhere to that treatment allocation are considered as having had zero MUSE sessions. For individuals who drop out from the (MUSE) treatment, all MUSE sessions before drop-out will be fully counted in. Individuals who have dropped out entirely from the study (including withdrawal from assessment) will be excluded from this analysis.

We will also compute the effect of (allocation and compliance to) MUSE on the number of CBT sessions required.

Additionally, an exploratory complier average causal effect (CACE) will be computed to estimate the effect that MUSE would likely have had if all participants had been compliant. Per-protocol or As-treated analyses are not envisaged but can be added in discussion with TSC if it turns out that treatment allocation adherence has been unexpectedly poor (non-adherence exceeding 10% of randomized individuals in any arm).

We will also report *therapy fidelity*, which is rated from therapy tapes from a 10% sample using a standard operating procedure. The acceptability scores for this will be presented as descriptive data and the results relate to the progression criteria (section 11.2).

Summaries of the modules used in the MUSE treatment will be reported to show the extent of the MUSE package utilised.

10.9 Additional unblinded analyses

We will carry out descriptive analyses, according to the same principles as laid out in Section 10.1, for the additional secondary outcomes from Section 8.8, and the additional feasibility outcomes from Section 8.9. This will include, among others, descriptive analysis of adverse events as well as satisfaction with therapy (STTS-R) by treatment group.

The interventions received in the TAU allocation group will be summarised to describe what was received in this treatment arm (e.g., needs based emotional support, psychoeducation, normalisation and stress management, formal CBT; information available from therapy packs or CSRI; see also Section 8.2). Additional TAU interventions received across both groups will be summarised according to treatment group (hospital services, community services, contacts with primary and community care professionals, criminal justice services).

Data concerning withdrawals and drop-outs, including the reasons for those, will be summarised by treatment arm.

Rates of discharge and reasons for discharge will be summarized according to treatment group.

Risks of bias across the trial outcomes will also be assessed according to the Cochrane Risk of Bias Tool (Higgins et al., 2019).

10.10. Protocol Deviations

Protocol deviations will be recorded and include issues like delayed randomisations, delay to provision of therapy by a Trial therapist, delayed follow up assessments, missing outcome assessments etc. These will be reported descriptively and not subject to formal analysis.

10.11. Timing, reporting and blinding

Recruitment for the trial started in May 2023 (actually a few days earlier, but any recruitments during the last days of April were added to the data for May), and was originally devised for 9 months, that is until end of January 2024. However, in their meeting on 26th July 2023, TMG agreed to a contingency option to extend the recruitment period until February 2024, which will be confirmed in January 2024. This means that complete baseline and post-treatment data will be available in the end of May 2024.

Data cleaning and timepoint data lock will be completed prior to data being provided to the statistical analysis team. The databases will be separated into data from each timepoint (baseline/12 week post-treatment/20 week follow-up) to allow interim analysis of data from the timepoints as they are completed.

Interim analysis on Sections 10.1 to 10.7 will begin as soon as the respective data are available, that is interim analysis of the baseline data will be carried out in March 2024, and interim analysis of the baseline/post-treatment data in June 2024. The reason to carry out these interim analyses is that the window for data analysis after the follow-up, within the funded grant period, is very short. Beginning with the interim analysis at these earlier stages allows an earlier understanding of the data, and the early production of code to carry out the final analysis. Statisticians will be entirely blinded to treatment allocation

during any interim analyses, and the results will not be unblinded for reporting purposes. In other words, any results communicated to TMG, CI or any other parties during the interim analysis phase will be in blinded form, and it will not be possible for any of these parties to unblind the treatment allocation.

The interim analysis should only begin after this Statistical Analysis Plan is approved by the Chief Investigator, trial statisticians and the Chair of the TSC.

The complete data including follow-up should be available by the end of July 2024. The final analysis will be performed at this point. After completion of the analysis for Sections 10.1 to 10.7 (which we will do in this anticipated order), the results from the blinded analysis will be reported to TMG and CI in a dedicated meeting. Following this meeting, the Statisticians will be unblinded, and the results can be used and communicated in unblinded form by all trial parties from this point.

The Analysis staff will then proceed with producing the final (unblinded) analysis for Sections 10.8 and 10.9.

Table 4. Timelines for Analysis and Unblinding (SAT= Statistical Analysis Team)

Time point	**Status**	**Action**
February 2024 (End)	Baseline data lock and cleaning;	Locked, cleaned, and blinded baseline data to be provided to SAT
March 2024	Descriptive Analysis of Baseline data (Section 10.1)	Blinded data analysis by SAT
May 2024 (End)	Post-treatment data lock and cleaning	Locked, cleaned, and blinded post-treatment data to be provided to SAT

June 2024	Descriptive Analysis of Follow-up data (Section 10.1) Difference-in-difference and model-based analyses according to Sections 10.2, 10.3, 10.5, 10.6	Blinded data analysis by SAT
July 2024 (End)	Follow-up data lock and cleaning	Locked, cleaned, and blinded follow-up data to be provided to SAT
August 2024	Descriptive Analysis of Follow-up data (Section 10.1) Difference-in-difference and model-based analyses according to Sections 10.2, 10.3, 10.4, 10.5, 10.6, 10.7	Blinded data analysis by SAT
August 2024 (End)	Reveal meeting	Presentation by SAT
September 2024	Adherence to treatment analyses and further descriptive analyses (Section 10.8, 10.9)	Unblinded data analysis by SAT
September 2024 (End)	Draft report delivery	Report by SAT

11. Progression and planning for definitive trial

Our progression criteria will follow signal of efficacy and cover domains of research delivery, therapy engagement and fidelity, and safety. These criteria have been developed

with LEAP to help determine if a full trial is warranted. A traffic-light system (green, amber, red) will be used to operationalise the progression criteria. We will use qualitative data to contextualise our progression criteria, to ensure that the participant feedback informs our understanding of our research delivery and signal of efficacy.

11.1 Signal of efficacy

i) Go: primary outcome data suggest the intervention may show an effect indicating clinical value warranting further investigation;

ii) Refine: primary outcome data indicate no measure of effect, but one or more secondary outcomes indicates an effect;

iii) Stop: no effect across any outcomes. Qualitative data will inform our understanding of any potential signal of efficacy, including whether the potential benefits of MUSE, for example, helping people understand the likely causes of unusual sensory experiences, are important to service users.

11.2 Progression criteria

The following are the approved trial monitoring and progression criteria thresholds for outcomes to determine progression to a future definitive trial. We focus here on the critical feasibility outcomes; detailed descriptions of further (secondary) feasibility and acceptability criteria relevant to the criteria are provided in a separate document produced in collaboration with the LEAP and TSC, which can be requested from the Trial manager.

Table 5: Approved trial monitoring and progression criteria to a future definitive trial

Progression criterion	**GREEN** (Feasibility of future trial demonstrated)	**AMBER** (Future trial will be feasible subject to identification of additional strategies to meet targets or	**RED** (Feasibility of future trial not demonstrated)	Relevant data obtained from this SAP in....

		remove barriers)		
Recruitment (the number of participants consented into the trial and randomised)	An average of least 7.84 participants are recruited and randomised per month (80% of recruitment target met).	At least 5.88 participants are recruited per month (60%-80% of recruitment target met).	An average of under 5.88 participant is recruited per month (under 60% of recruitment target met).	Section 10.1
Therapy engagement (% who drop-out of therapy)	At least 80% of the participants in the intervention arm completed at least 4 out of the 6-8 sessions of MUSE.	If 60-80% of participants in the intervention arm complete at least 4 out of the 6-8 sessions of MUSE.	If less than 60% of participants in the intervention arm complete at least 4 out of the 6-8 sessions of MUSE.	Section 10.8
Assessment retention (% of participants who are *not* lost to follow-up at primary assessment endpoint (12 weeks post randomisation)	At least 70% of participants complete primary outcome measure at primary assessment endpoint.	50-70% of participants complete primary outcome measure at primary assessment endpoint.	Less than 50% of participants complete primary outcome measure at primary assessment endpoint.	Section 10.1

Therapy fidelity (Adherence ratings from therapy tapes)	Over 80% of rated therapy tapes are rated as acceptable.	50-80% of rated therapy tapes are rated as acceptable.	Less than 50% of rated therapy tapes will be rates as acceptable.	Section 10.8
Safety (number of related SAEs)	0-1 Related SAEs in the Intervention arm.	2 Related SAEs in the Intervention arm.	3+ Related SAEs in the Intervention arm.	Section 10.9, 12

11.3 Sample sizes for a definitive trial

Using the estimates from Sections 10.1 to 10.4 but considering the primary outcomes only (and for PYSRATS only the Total AH score), we will estimate the minimum sample size for a definitive trial to detect an effect of this estimated size at a type-1 error rate of 5% with 80% power. Under the assumption that the primary outcomes have equal true effect sizes, this will constitute an asymptotically unbiased estimate of the minimum sample size necessary to detect this common effect. These sample size calculations will also require estimates of the standard deviation of the respective outcomes. For each of the outcome measures, this will be calculated as the pooled standard deviation across the two treatment arms. Should the total sample size available for this purpose be less than 70, an appropriate inflation factor will be applied to the pooled standard deviation, yielding a minimum required sample size for the respective outcome measure to ensure a given power of the definitive trial with 80% confidence (Teare et al., 2014).

12. Safety Analyses

All analyses will be in line with the safety reporting and safety criteria in the protocol. Serious Adverse Events (Adverse Events which meet the criteria for seriousness), and Adverse Events of interest as per the Protocol will be captured for the participants. AEs and SAEs will be tabulated per trial arms and the action taken, outcome, relatedness and expectedness in the opinion of the investigator will be reported using frequency tables. Any Urgent Safety Measures will be reported, along with what measures were taken in

accordance with the Protocol and advice from Sponsor and REC, if there have been any such measures.

In case of a red number of SAEs according to Sec 11.2., a sensitivity analysis for the primary outcomes using a safety population (all subjects who received any study treatment incl. control which are not dropped out due to serious adverse events post randomisation) will be added. In case of an amber number of SAEs, it will be discussed with TSC whether such an analysis is deemed necessary and carried out if so.

13 Reporting conventions and principles

Means, standard deviation, and any other statistics other than quantiles, will be reported to two decimal places greater than the original data. Quantiles, such as median, or minimum and maximum will use the same number of decimal places as the original data. Estimated parameters not on the same scale as raw observations (e.g. regression coefficients and confidence intervals) will be reported to 2 decimal places.

The population to be used will generally be ITT unless for specific purposes as outlined earlier in this SAP. In any case, the studied population will be explicitly set at the start of the file or block of code that computes the output.

Any code will have

- The date and time included
- The name of the code file that produced the analysis
- The author
- The date and time of writing
- References to inputs and outputs
- Reference to any parent code file that runs the child code file.

References

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire – Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, *65*, 48-58. https://doi.org/https://doi.org/10.1016/j.concog.2018.07.001

Andrade, J., May, J., Deeprose, C., Baugh, S. J., & Ganis, G. (2014). Assessing vividness of mental imagery: The Plymouth Sensory Imagery Questionnaire. *Br J Psychol*, *105*(4), 547-563. https://doi.org/10.1111/bjop.12050

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med*, *2*(4), 297-307. https://doi.org/10.1016/s1389-9457(00)00065-4

Bosnjak Kuharic, D., Kekin, I., Hew, J., Rojnic Kuzman, M., & Puljak, L. (2019). Interventions for prodromal stage of psychosis. *Cochrane Database Syst Rev*, *2019*(11). https://doi.org/10.1002/14651858.CD012236.pub2

Cloitre, M., Shevlin, M., Brewin, C. R., Bisson, J. I., Roberts, N. P., Maercker, A., Karatzias, T., & Hyland, P. (2018). The International Trauma Questionnaire: development of a self-report measure of ICD-11 PTSD and complex PTSD. *Acta Psychiatr Scand*, *138*(6), 536-546. https://doi.org/10.1111/acps.12956

Cooper, M., Messow, C. M., McConnachie, A., Freire, E., Elliott, R., Heard, D., ... & Morrison, J. (2018). Patient preference as a predictor of outcomes in a pilot trial of person-centred counselling versus low-intensity cognitive behavioural therapy for persistent sub-threshold and mild depression. *Counselling Psychology Quarterly*, 31(4), 460-476.

Dalenberg, C., and E. Carlson. (2010). "Severity of Dissociative Symptoms-Adult (Brief Dissociative Experiences Scale (DES-B)-Modified)." Washington, DC, American Psychiatric Association.

Dahmen, G., & Ziegler, A. (2004). Generalized estimating equations in controlled clinical trials: hypotheses testing. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 46(2), 214-232.

Dodgson, G., Aynsworth, C., Mitrenga, K. J., Gibbs, C., Patton, V., Fernyhough, C., Dudley, R., Ewels, C., Leach, L., Alderson-Day, B., & Common, S. (2021). Managing unusual sensory experiences: A feasibility trial in an At Risk Mental States for psychosis group. *Psychology and Psychotherapy: Theory, Research and Practice*, *94*(3), 481-503. https://doi.org/https://doi.org/10.1111/papt.12323

Dodgson, G., Alderson-Day, B., Smailes, D., Ryles, F., Mayer, C., Glen-Davison, J., ... & Fernyhough, C. (2021b). Tailoring cognitive behavioural therapy to subtypes of voice-hearing using a novel tabletised manual: a feasibility study. *Behavioural and Cognitive Psychotherapy*, *49*(3), 287-301.

Dudley, R., Dodgson, G., Common, S., O'Grady, L., Watson, F., Gibbs, C., Arnott, B., Fernyhough, C., Alderson-Day, B., Ogundimu, E., Kharatikoopaei, E., Patton, V., & Aynsworth, C. (2022). Managing Unusual Sensory Experiences in People with First-Episode Psychosis (MUSE FEP): a study protocol for a single-blind parallel-group randomised controlled feasibility trial. *BMJ Open*, *12*(5), e061827. https://doi.org/10.1136/bmjopen-2022-061827

Dudley et al (202x). In preparation.

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. BMJ, 355.

Fernyhough, C., Bland, K., Meins, E., & Coltheart, M. (2007). Imaginary companions and young children's responses to ambiguous auditory stimuli: implications for typical and atypical development. *J Child Psychol Psychiatry*, *48*(11), 1094-1101. https://doi.org/10.1111/j.1469-7610.2007.01789.x

Fernyhough, C., Watson, A., Bernini, M., Moseley, P., & Alderson-Day, B. (2019). Imaginary Companions, Inner Speech, and Auditory Verbal Hallucinations: What Are the Relations? [Original Research]. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01665

Freeman, D., Bradley, J., Waite, F., Sheaves, B., DeWeever, N., Bourke, E., McInerney, J., Evans, N., Černis, E., Lister, R., Garety, P., & Dunn, G. (2016). Targeting Recovery in Persistent Persecutory Delusions: A Proof of Principle Study of a New Translational Psychological Treatment (the Feeling Safe Programme). *Behav Cogn Psychother*, *44*(5), 539-552. https://doi.org/10.1017/s1352465816000060

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L. M., Nickless, A., Harrison, P. J., ... & Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. *The Lancet Psychiatry*, 4(10), 749-758.

Gagnon, C., Bélanger, L., Ivers, H., & Morin, C. M. (2013). Validation of the Insomnia Severity Index in primary care. *J Am Board Fam Med*, *26*(6), 701-710. https://doi.org/10.3122/jabfm.2013.06.130064

Goldman, H. H., Skodol, A. E., & Lave, T. R. (1992). Revising axis V for DSM-IV: a review of measures of social functioning. *Am J Psychiatry*, *149*(9), 1148-1156. https://doi.org/10.1176/ajp.149.9.1148

Haddock, G., McCarron, J., Tarrier, N., & Faragher, E. B. (1999). Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychol Med*, *29*(4), 879-889. https://doi.org/10.1017/s0033291799008661

Haselgruber, A., Sölva, K., & Lueger-Schuster, B. (2020). Symptom structure of ICD-11 Complex Posttraumatic Stress Disorder (CPTSD) in trauma-exposed foster children: examining the International Trauma Questionnaire - Child and Adolescent Version (ITQ-CA). *Eur J Psychotraumatol*, *11*(1), 1818974. https://doi.org/10.1080/20008198.2020.1818974

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions* (J. P. T. Higgins & S. Green, Eds. Vol. 4). John Wiley & Sons.

Higgins, J. P., Savović, J., Page, M. J., Elbers, R. G., & Sterne, J. A. (2019). Assessing risk of bias in a randomized trial. *Cochrane handbook for systematic reviews of interventions*, 205-228.

Keetharuth, A. D., Brazier, J., Connell, J., Bjorner, J. B., Carlton, J., Taylor Buck, E., Ricketts, T., McKendrick, K., Browne, J., Croudace, T., & Barkham, M. (2018). Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *Br J Psychiatry*, *212*(1), 42-49. https://doi.org/10.1192/bjp.2017.10

Keetharuth, A. D., Bjorner, J. B., Barkham, M., Browne, J., Croudace, T., & Brazier, J. (2021). An item response theory analysis of an item pool for the recovering quality of life (ReQoL) measure. *Qual Life Res*, *30*(1), 267-276. https://doi.org/10.1007/s11136-020-02622-2

Keetharuth, A. D., Rowen, D., Bjorner, J. B., & Brazier, J. (2021b). Estimating a preference-based index for mental health from the recovering quality of life measure: valuation of recovering quality of life utility index. *Value in Health*, *24*(2), 281-290.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*, *16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kuznetsova, O. M., & Tymofyeyev, Y. (2012). Preserving the allocation ratio at every allocation with biased coin randomisation and minimization in studies with unequal allocation. *Statistics in Medicine*, *31*(8), 701-723. https://doi.org/https://doi.org/10.1002/sim.4447

Miller, Brian J et al. (2019). Insomnia and suicidal ideation in nonaffective psychosis. *Sleep,* 42,2 (2019): zsy215. doi:10.1093/sleep/zsy215

Moritz, S., Gawęda, Ł., Heinz, A., & Gallinat, J. (2019). Four reasons why early detection centers for psychosis should be renamed and their treatment targets reconsidered: we should not catastrophize a future we can neither reliably predict nor change. *Psychol Med*, *49*(13), 2134-2140. https://doi.org/10.1017/s0033291719001740

Moseley, P., Aleman, A., Allen, P., Bell, V., Bless, J., Bortolon, C., Cella, M., Garrison, J., Hugdahl, K., Kozáková, E., Larøi, F., Moffatt, J., Say, N., Smailes, D., Suzuki, M., Toh, W. L., Woodward, T., Zaytseva, Y., Rossell, S., & Fernyhough, C. (2021). Correlates of Hallucinatory Experiences in the General Population: An International Multisite Replication Study. *Psychological Science*, *32*(7), 1024-1037. https://doi.org/10.1177/0956797620985832

Paulik, G., Badcock, J. C., & Maybery, M. T. (2007). Poor intentional inhibition in individuals predisposed to hallucinations. *Cognitive Neuropsychiatry*, *12*, 457-470. https://doi.org/10.1080/13546800701394329

Pincus, T., Miles, C., Froud, R., Underwood, M., Carnes, D., & Taylor, S. J. (2011). Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. *BMC Medical Research Methodology*, 11, 1-14.

Redican, E., Nolan, E., Hyland, P., Cloitre, M., McBride, O., Karatzias, T., Murphy, J., & Shevlin, M. (2021). A systematic literature review of factor analytic and mixture models of ICD-11 PTSD and CPTSD using the International Trauma Questionnaire. *J Anxiety Disord*, *79*, 102381. https://doi.org/10.1016/j.janxdis.2021.102381

Sackett, D. L. (1979). Bias in analytic research. In *The case-control study consensus and controversy* (pp. 51–63). Elsevier.

Smailes, D., Burdis, E., Gregoriou, C., Fenton, B., & Dudley, R. (2020). Pareidolia-proneness, reality discrimination errors, and visual hallucination-like experiences in a non-clinical sample. *Cognitive Neuropsychiatry*, *25*, 113-125. https://doi.org/10.1080/13546805.2019.1700789

Spielberger, C. D. (1983). State-Trait Anxiety Inventory. In *The Corsini Encyclopedia of Psychology*. https://doi.org/https://doi.org/10.1002/9780470479216.corpsy0943

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*, *166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Teare, M. Dawn, et al. (2014). "Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study." Trials 15, 1-13.

Williams, R., Farquharson, L., Palmer, L., Bassett, P., Clarke, J., Clark, D. M., & Crawford, M. J. (2016). Patient preference in psychological treatment and associations with self-reported outcome: national cross-sectional survey in England and Wales. *BMC psychiatry*, 16(1), 1-8.

Woodward, T. S., Jung, K., Hwang, H., Yin, J., Taylor, L., Menon, M., Peters, E., Kuipers, E., Waters, F., Lecomte, T., Sommer, I. E., Daalman, K., van Lutterveld, R., Hubl, D., Kindler, J., Homan, P., Badcock, J. C., Chhabra, S., Cella, M., Keedy, S., Allen, P., Mechelli, A., Preti, A., Siddi, S., & Erickson, D. (2014). Symptom dimensions of the psychotic symptom rating scales in psychosis: a multisite study. *Schizophr Bull*, *40 Suppl 4*(Suppl 4), S265-274. https://doi.org/10.1093/schbul/sbu014

Wykes, T., Bell, A., Carr, S., Coldham, T., Gilbody, S., Hotopf, M., Johnson, S., Kabir, T., Pinfold, V., Sweeney, A., Jones, P. B., & Creswell, C. (2021). Shared goals for mental health research: what, why and when for the 2020s. *Journal of Mental Health*, 1-9. https://doi.org/10.1080/09638237.2021.1898552

Yung, A. R., Yuen, H. P., McGorry, P. D., Phillips, L. J., Kelly, D., Dell'Olio, M., Francey, S. M., Cosgrave, E. M., Killackey, E., Stanford, C., Godfrey, K., & Buckby, J. (2005). Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. *Aust. N. Z. J. Psychiatry 39*, 964–971. https://doi.org/10.1080/j.1440-1614.2005.01714.x. PMID: 16343296.

Appendix: Consort guidelines checklist

In this appendix, we report concordance with the CONSORT guidelines for publication of a pilot trial (Eldridge et al., 2016). Not all elements are relevant to this SAP, and we have marked irrelevant elements as 'NA'.

Section/topic	**Num.**	**Extension for pilot trials**	**Section**
Title/abstract	1a	Identification as a pilot or feasibility randomised trial in the title	Front matter
	1b	Structured summary of pilot trial design, methods, results, and conclusions (for specific guidance see CONSORT abstract extension for pilot trials)	6, 8
Introduction	2a	Scientific background and explanation of rationale for future definitive trial, and reasons for randomised pilot trial	4.1, 4.2
	2b	Specific objectives or research questions for pilot trial	5
Trial design	3a	Description of pilot trial design (such as parallel, factorial) including allocation ratio	6
	3b	Important changes to methods after pilot trial commencement (such as eligibility criteria), with reasons	6
Participants	4c	How participants were identified and consented	Protocol
Outcomes	6a	Completely defined prespecified assessments or measurements to address each pilot trial objective specified in 2b, including how and when they were assessed	8

	6b	Any changes to pilot trial assessments or measurements after the pilot trial commenced, with reasons	NA
	6c	If applicable, prespecified criteria used to judge whether, or how, to proceed with future definitive trial	11.2
Sample size	7a	Rationale for the numbers in the pilot trial	9, 11
Sequence generation	8b	Type of randomisation(s); details of any restriction (such as blocking and block size)	7.1
Analytical methods	12a	Methods used to address each pilot trial objective whether qualitative or quantitative	8,10
Participant flow	13a	For each group, the numbers of participants who were approached and/or assessed for eligibility, randomly assigned, received intended treatment, and were assessed for each objective	8.3
Recruitment	14b	Why the pilot trial ended or was stopped	9, 10.11
Numbers analysed	16	For each objective, number of participants (denominator) included in each analysis. If relevant, these numbers should be by randomised group	10.1
Outcomes and estimation	17a	For each objective, results including expressions of uncertainty (such as 95% confidence interval) for any estimates. If relevant, these results should be by randomised group	10
Ancillary analyses	18	Results of any other analyses performed that could be used to inform the future definitive trial	10, 11

Harms	19a	If relevant, other important unintended consequences	12
Limitations	20	Pilot trial limitations, addressing sources of potential bias and remaining uncertainty about feasibility	6.3, 10.9
Generalisability	21	Generalisability (applicability) of pilot trial methods and findings to future definitive trial and other studies	4, 11
Interpretation	22	Interpretation consistent with pilot trial objectives and findings, balancing potential benefits and harms, and considering other relevant evidence	NA
	22a	Implications for progression from pilot to future definitive trial, including any proposed amendments	11
Registration	23	Registration number for pilot trial and name of trial registry	Front matter
Protocol	24	Where the pilot trial protocol can be accessed, if available	Front matter
Funding	26	Ethical approval or approval by research review committee, confirmed with reference number	Front matter