Randomised controlled feasibility trial of an integrated psychosexual intervention for sexual difficulties in people with Multiple Sclerosis: the PIMS trial

Statistical Analysis Plan

Version: 1.0

Date: 30/07/2024

Author: Sam Norton

Chief Investigator: Prof Rona Moss-Morris

Signature: _____ Date: 09/07/2026

Senior Statistician: Dr Sam Norton

Signature: Date: 30/07/2024

Trial Registration: ISRCTN 12202900

Document version history log:

Date:	Version:	Changes made:
09/01/2024	0.1	Beta version
17/01/2024	0.2	Minor updates to outcomes and analyses
30/07/2004	1.0	Queries resolved and set to V1

1 CONTENTS

2 DESCRIPTION OF THE TRIAL

2.1 STUDY AIMS & OBJECTIVES

PIMS is a two-armed, parallel group, multi-centre randomised controlled trial (RCT) designed to assess the feasibility of conducting a full-scale (phase III) definitive trial of the effectiveness of the PIMS intervention in people with MS experiencing sexual or intimacy difficulties.

The primary objective is to evaluate the feasibility of a larger randomised controlled trial based on eligibility and recruitment rates, retention, and adherence. This is evaluated based on the progression criteria (see section 4.2.1), which is a composite of the feasibility objectives:

1. Determine eligibility rates (proportion of those screened who meet inclusion criteria)
2. Recruitment rates (proportion of eligible PwMS consented to trial)
3. Assess retention of participants (% participants who complete 14-week follow up assessment)
4. Assess participant adherence
    a. Average number of appointments attended
    b. Average number of self-guided sessions completed
5. Acceptability as indicated by:
    a. Average acceptability rating of the intervention
    b. Average score on 8-item satisfaction inventory
    c. Acceptability of self-reported outcome measures for a future trial (i.e., completion rates, reliability, item-level missing data, floor/ceiling effects)

Other feasibility objectives include determining the following for research sites or HCPs

1. Recruitment rates
    a. Number of NHS sites recruiting for trial
    b. Number of facilitators recruited from neurology services
2. Facilitator adherence
    a. Proportion of required session material covered during appointments

Secondary objectives include

1) To explore the differences in means on self-reported outcomes between trial arms and at between pre-post randomisation time points
2) To qualitatively explore the participant perceptions of the acceptability and usefulness of the PIMS/PSE packages and identify areas of improvement for a full-scale trial
3) To qualitatively explore HCPs perceptions of acceptability and usefulness of the PIMS/PSE packages and barriers/facilitators to delivery
4) To explore the feasibility of collecting health economic estimates of changes in service use associated with PIMS and sensitive measures of quality of life (QoL)

The analyses described in this document relate to the evaluation the feasibility objectives and the first of the secondary objectives to explore mean differences of the PIMS intervention compared to TAU and will be included in the main study report.

Details of other secondary analyses, including analyses of data form nested qualitative study and exploratory quantitative analysis of the health economic data, are not covered explicitly in this document. Relevant information regarding these will be added to the appendix.

2.2 Design, Randomization & Blinding

The PIMS trial is a two-armed, parallel group, multi-centre RCT. Participants are individually randomised following a 1:1 ratio to receive either the PIMS intervention or TAU. Follow up assessment takes place at 14-weeks post-randomisation for all randomised participants. The recruitment and data collection procedure is shown in Figure 1.

Random allocation to the two groups uses random block sizes and is stratified by recruiting site and gender identity, implemented via the King's Clinical Trials Unit online system. Due to the anticipated low number of people recruited that identify as transgender/non-binary these people are all allocated to the treatment arm.

Blinding of the participants is not possible given the nature of the intervention. The therapist and researcher undertaking the randomisation and providing patients with support during the intervention will also not be blind to treatment allocation.

The statistician will remain blinded to treatment allocation until after the completion of the analyses described in this report. Unblinding of the statistician will occur prior to undertaking any analysis of the adherence to the intervention sessions, since these analyses are exploratory they may be undertaken by an unblinded researcher involved in other elements of the trial.

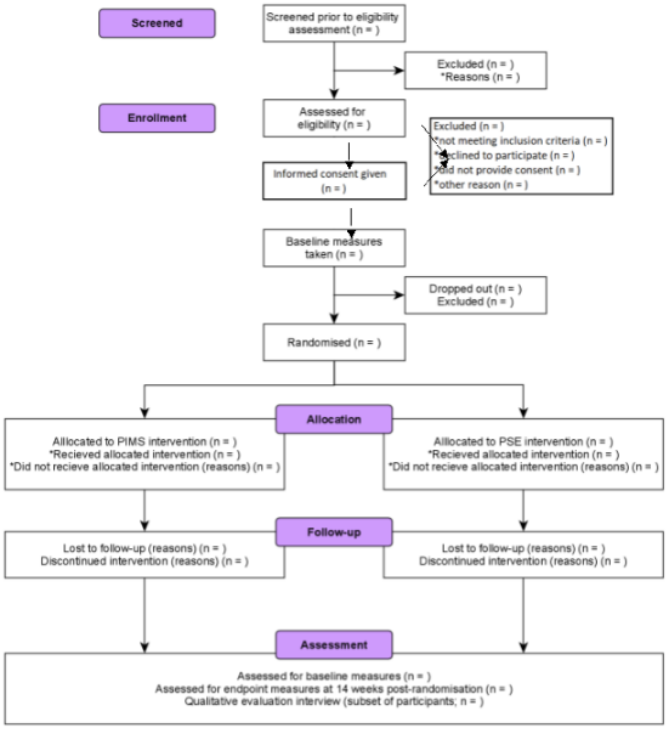


Figure 1. Trial design flow diagram

2.3 ELIGIBILITY CRITERIA

The specific inclusion and exclusion for the trial are listed below.

Inclusion criteria: Confirmed diagnosis of Multiple Sclerosis (using McDonalds revised criteria); Attending the Neurology clinic at one of the participating research sites; Aged ≥ 18; Affirmative response to the following question: Are you currently experiencing any sort of sexual or intimacy difficulty?; Able and willing to provide informed consent.

Exclusion criteria: Severe visual or hearing impairments which would impact the ability to engage in the intervention; Severe mental health disorder (e.g., psychosis) which would impact the ability to engage in the intervention; Currently receiving psychotherapy for sex or intimacy difficulties or are participating in any other psychological intervention trial; Insufficient verbal and/or written proficiency in English.

For facilitators participating in nested qualitative study, the inclusion criteria are: Clinical staff at research sites that work with people with Multiple Sclerosis. Exclusion criteria are: unable or unwilling to consent to participate in the follow up interviews.

2.4 SAMPLE SIZE CALCULATION

A sample of 50 is within the recommended sample size for feasibility trials (Teare et al., 2014). As significance testing is not used in trials of this nature, the sample size is based on the precision with which the key parameters are estimated. A sample size of 50 allows us to estimate recruitment rates with a 95%CI of width +/-8% assuming 50% recruitment (width will be narrower if recruitment lower) and drop-out rates with a 95%CI with maximum width ±14% overall and ±20% within each group. A sample size of 50 overall also provides an estimate of the standard deviation in the study population that is sufficiently accurate to be used in the sample size estimation for a definitive trial.

2.5 OUTCOME MEASURES

2.5.1 Primary Feasibility Parameters

Primary feasibility parameters will be assessed by collecting descriptive data on recruitment and retention rates and willingness to be randomised according to Consolidated Standards of Reporting Trials (CONSORT) feasibility and pilot trial guidelines [50]. The outcomes will be collected by someone separate to the facilitator. We will determine the following:

- Eligibility and recruitment rates: This will allow us to determine whether our eligibility criteria are appropriate, and whether those eligible to participate (i.e., complete screening) are willing to be randomised.
    - Eligibility rate: proportion of those screened who meet inclusion criteria
    - Recruitment rate: proportion of eligible PwMS consented to trial
- Follow up/retention rates: Proportion of participants who were randomised that completed the follow up assessment
    - Retention rate: % participants randomised who complete 14-week follow up
- Treatment adherence rates
    - Total number of appointments attended/completed
    - Number of appointments attended with facilitator

    - Number of self-guided sessions completed: at least on activity completed for a session will be considered partial completion for the session
    - Number of activities completed by session
    - Adherence defined as at least one session with a facilitator and at least 3 out of 6 self-guided sessions
- Acceptability: This will be evaluated based on theoretical framework for acceptability that have been developed into an 8-item scale for use in feasibility trials [51-53]. Items are on 5-point Likert response scales, though scale point labels differ based on the item. Wording of the items have been adapted for this study, as is directed by the scale authors (see Appendix A).

    - Average acceptability ratings for the TFA Scale items in the intervention group

2.5.2 Secondary feasibility parameter

Other feasibility objectives include determining the following for research sites or HCPs

- Recruitment rates
    - Number of NHS sites recruiting for trial
    - Number of facilitators recruited from neurology services
- Facilitator adherence
    - Proportion of required session material covered during appointments

2.5.3 Patient reported outcomes and other data collected

Secondary outcome measures will be used to identify a series of well-established endpoints of clinical importance. These will be investigated for their suitability for use as a primary outcome measure in a future trial. Basic psychometric properties of the scales will be considered to determine appropriateness of self-reported outcome measures for a future trial will (i.e., completion rates, reliability, item-level missing data, floor/ceiling effects, scale validity ratings)

- **Sex- and intimacy-related outcomes:**
    - Sexual Function Evaluation Questionnaire (SFEQ). This scale includes a total of 16-items formed into four subscales of four items each, which are calculated as a weighted sum score based recommended weights. Items scores range from 0 to 4 using a hurdle approach based on a mix of binary and ordinal responses (0 means absence and 1 to 4 indicated presence and severity). A total score is also generated based on the sum scores of the four subscales and will be used in favour of the subscale to reduce the overall number of outcomes reported. The specific variables used in the analysis will be:
        - SFEQ total score
    - Multiple Sclerosis Intimacy and Sexuality Questionnaire (MSISQ-19). This scale includes a total of 19 items forming three subscales. Each item is scored on a 1 to 5 ordinal frequency scale (never to always). The specific variables used in the analysis will be:
        - MSISQ-19 primary score
        - MSISQ-19 secondary score
        - MSISQ-19 tertiary score
    - Female Sexual Distress Scale-Revised (FSDS-R). This scale includes 13 items summed to form a total score. Item responses are ordinal frequency scale from 0 to 4 (never to always). The specific variables used in the analysis will be:

Commented [SN1]: Check there's no total score

    - FSDS-R total score
  - New Sexual Satisfaction Scale – Short form (NSSS-S). This scale includes 12 items summed to form a total score. Item responses are ordinal likert type scale from 1 to 5 (strongly agree to strongly disagree). The specific variables used in the analysis will be:
    - NSSS-S total score
- **Assessment of disability and QoL:**
  - Guy's Neurological Disability Scale (GNDS).
    - GNDS total score
    - GNDS sexual score
  - EQ-5D-5L. This five item scale includes assessments of domains of quality of life on a five point ordinal response scale. A utility score is calculated based on a cross-walk using UK population derived values sets.
    - EQ-5D-5L utility score
    - EQ-5D-5L health score
  - Patient Health Questionnaire-9 (PHQ-9). This 9 item questionnaire
    - PHQ9 total score
  - Generalised Anxiety Disorders-7 questionnaire (GAD-7).
    - GAD7 total score
- **Process outcomes:**
  - Five Facet Mindfulness Questionnaire (FFMQ)
    - FFMQ awareness of actions score
    - FFMQ non-judgement score
  - Dyadic Sexual Communication Scale (DSC). This scale includes 13 items summed to form a total score. Item responses are ordinal likert type scale from 1 to 5 (strongly agree to strongly disagree). The specific variables used in the analysis will be:
    - DSC total score

Commented [SN2]: Confirm only these two subscales are used

Commented [SN3R2]: Also whether the full or shortened version

3 STATISTICAL CONSIDERATION

3.1 TIMING

The main analysis will be performed only once the database has been cleaned and locked and after the finalisation and approval of the statistical analysis plan. Results will be presented as part of a report to the TSC prior to unblinding. This report will be partially blinded with the groups referred to as Arm 1 and ARM 2.

3.2 ANALYSIS SETS

For the analysis, the sample will be described in terms of the *intention-to-treat sample* and the *per-protocol sample*. For each sample, the analysis will follow the intention-to-treat principle with participants analysed in the treatment groups to which they are randomised.

Intention-to-treat sample: All participants who complete the baseline assessment, are randomised, and complete at least one post-treatment assessment will be included in this sample. Participants who withdraw from the study after randomisation will have their data included up to the point of withdrawal. A strict intention-to-treat analysis including all

participants may not be possible since those withdrawing prior to completing any follow-up assessments cannot be included in the analysis.

Per-protocol sample: All participants in the intention-to-treat who do not deviate substantially from the protocol and will be included in the per-protocol sample. Exclusion will be determined on a per-participant basis according to the following principles: i) for those in the intervention group, where the treatment dose received by the participant is minimal, which is operationalised as undertaking at least one facilitated session plus at least one self-directed session; and ii) from the control group where the participant is known to have received some form of psychoeducational intervention, including PIMS inadvertently.

Commented [SN4]: Need to decide

Commented [SN5R4]: Any facilitated session attended plus at least one self-directed session. Check amber progression criteria

3.3 Statistical software

Participant self-report forms are collected online using RedCap. The postdoctoral researcher managing data collection will extract data as needed and provide these in a Stata .dta format, which includes coded value labels.

Stata 18.0 or higher will be used for the descriptive feasibility and exploratory efficacy analyses.

3.4 Reporting conventions

Counts will be reported to zero decimal places. Percentages will be reported to one decimal place. The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place 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 effect sizes) will be reported to 3 significant figures. Significance levels (i.e., *p*-values) will be reported to 3 decimal places; with *p*-values less than 0.001 reported as $p<0.001$.

3.5 Practical measures to minimise bias

No interim efficacy analysis is to be conducted, which mitigates any potential for the information to influence any of those involved in the conduct of the trial. Prior to the main analysis being conducted the database will be 'locked' and no further amendments to the study data allowed. The system used for data collection does not include a feature to 'lock' the database so this will involve withdrawing the ability for additional participant data to be added and a final data extract being passed to the statistician in its original format.

The main analysis will be conducted by the trial statistician who will be unblinded only once the main analysis has been completed. Partially blinded group allocations will be obtained by the statistician from the KCTU randomisation service. The main efficacy report will be presented with groups indicated using numbers (i.e., arms "1" and "2") with the group allocations only revealed once the details of the report have been confirmed as correct.

Safety monitoring decisions will remain isolated from the efficacy information. Furthermore, adverse events will be coded by the chief investigator removing any identifiable data from the data extract (i.e., redacting IDs and any other information that might reveal group allocation).

3.6 SCORING OF QUESTIONNAIRES

For all scales (and subscales) the scoring is straightforward and based on calculating the total score across items forming the scale. Missing data for individual items within questionnaires will be accounted for via pro-rating, which is equivalent to mean imputation. Specifically, the total pro-rata score will be calculated using the following formula provided that the number of missing items is less than or equal to 33% of the total number of items for the scale.

$$Y = N_t * \frac{\sum x_n}{N_c}$$

Where *Y* is the pro-rata score for the scale, N_t is the total number of items included in the scale, x_n is the score for item *n*, and N_c is the number of items completed. This constitutes the mean score across completed items multiplied by the total number of items. For example, for a scale involving 6 items, for which data are available for only 4 items with ratings of 4, 4, 5, 5, the pro-rata total score will be calculated as 6*((4+4+5+5)/4) = 27.

4 STATISTICAL ANALYSES

4.1 DESCRIPTIVE ANALYSES

All variables will be described by treatment group and assessment time point. In addition to tables showing descriptive statistics by treatment group and assessment time point, further tables will be presented with variables described by recruitment site. All continuous variables will be summarised using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, median, minimum, and maximum. The frequency and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical variables. All tables will be annotated with the total size of the analysis sample used to generate the table, including any missing observations.

4.2 PRIMARY ANALYSES

4.2.1 Primary outcome

The key analyses undertaken inform the decision to progress to a definitive trial based on the feasibility parameters and acceptability of the intervention. These are described in Section 2.5.1 as the primary outcomes. A decision to progress will be judged via a commonly used traffic light system shown below (Table 3). All parameters listed in criteria 1 to 5 will be reported descriptively as numerator/denominator and percentages with 95% CIs calculated using the binomial exact method.

- Green- feasible to proceed to a full RCT with the current trial design/procedures
- Amber- modifications are required before proceeding to a definitive trial
- Red- not feasible to proceed to a definitive trial.

Where all criteria are observed to be green, the recommendation will be that a full-scale trial is feasible without major modifications to the intended protocol (i.e., progress). Where there is a mixture of green and amber criteria, the recommendation will be that a full-scale trial is feasible only if modifications can be made to the relevant domain such that the green

judgment is likely to be achieved. It is anticipated that these modifications would then need to be confirmed as part of an internal pilot for the full-scale trial. Where any of the criteria are observed to be red, the recommendation will be that a full-scale trial is not feasible (i.e., do not progress) and that major modifications to the protocol are needed and a further feasibility study likely to be necessary.

Table 1. Progression criteria for full-scale trial.

Progression criteria	Judgement
1. Eligibility: Meeting eligibility estimates (PwMS)	Green: ≥33% of those screened meet eligibility criteria Amber: 10-32% of those screened meet eligibility criteria Red: <10% of those screened meet eligibility criteria
2. Recruitment: Meeting recruitment targets (sites and PwMS)	Green: Two or more research sites recruited & ≥33% of eligible PwMS consent to take part in the trial. Amber: Two or more research sites recruited & 5-32% of eligible PwMS consent to take part in the trial. Red: Fewer than two research sites recruited OR two or more research sites recruited & <5% of eligible PwMS consent to take part in the trial.
3. Retention: Assessments made with satisfactory follow-up rates (PwMS)	Green: ≥80% complete the 14-week post randomisation follow-up assessment Amber: 50-79% complete the 14-week post randomisation follow-up assessment Red: <50% complete the 14-week post randomisation follow-up assessment
4. Adherence: Completion of sufficient number of intervention sessions (PwMS)	Green: If ≥80% complete at least one session with a facilitator and at least 3 out of 6 self-guided sessions Amber: If 50-79% complete at least one session with a facilitator and at least 3 out of 6 self-guided sessions Red: If <50% complete at least one session with a facilitator and at least 3 out of 6 self-guided sessions
5. Satisfaction: Rating of overall satisfaction with intervention package (PwMS)	Green: If ≥80% of participants report satisfaction with the intervention package, i.e., select 'agree' or 'strongly agree' to the overall satisfaction question. (Participants who drop out and do not provide data will be assumed to not be satisfied) Amber: If 50-79% of participants report satisfaction with the intervention package Red: If <50% of participants report satisfaction with the intervention package

4.2.2 Patient reported outcomes

Adjusted between-group differences at the 14-week post-randomisation follow-up assessment will be reported with 95%CIs based on robust standard errors. These will be estimated using linear mixed-effect regression models, with a random effect to account for clustering by therapist in the intervention arm. Models will include treatment group as a dummy coded variable along with baseline level of the outcome and stratification factors.

Commented [SN6]: Protocol says mixed-effects models but we only have one follow up. Need to discuss whether we account for clustering within centre/therapist

Commented [SN7R6]: Rando effect for therapist

Given the study is not powered to determine superiority of the intervention versus control, *p*-values for between group differences will not be reported and interpretation will be based on

95%CIs indicating the range of plausible treatment effects supported by the data. Standard deviations for the primary outcomes will be computed to estimate the sample size for a definitive trial. Standardised effect sizes will be reported along with 95%Cis using the Hedge's g formula with small sample correction.

4.2.3 Model assumption checks

All regression models will be assessed for any important violation of the underlying assumptions. All models for continuous outcomes, assuming normally distributed and homoscedastic residuals, will be assessed using Q-Q plots for the residuals and scatterplots of fitted versus residual values. These will also be used to identify potential outliers and influential observations. To protect against minor deviations from homoscedasticity, robust standard errors will be estimated following the Huber-White method. Where substantial deviations occur transformations will be considered.

4.2.4 Missing data

The number (%) with complete data will be reported for all variables and clearly indicated in the descriptive tables. The total number (%) of people providing no post-randomisation information (that cannot be included in the analysis) will be presented by treatment group. Exploratory analysis will consider baseline variables related to attrition and summarised as odds ratios with 95%CIs.

In the treatment efficacy models, missing baseline covariates will be handled using the missing indicator method (White & Thompson, 2005). This involves imputing missing values with the mean of the observed values and including indicator variables for missingness for each variable with missing data in the model.

4.2.5 Sensitivity analyses

Sensitivity analyses regarding missing data are described above.

4.2.6 Interim analyses

Interim analyses considering eligibility and recruitment rates are undertaken throughout the trial. No interim analyses are planned to assess treatment efficacy.

4.3 PROCESS ANALYSES

4.3.1 Moderators

The study is not specifically powered to examine moderators and therefore moderation analysis will not be undertaken to assess treatment effect heterogeneity.

Commented [SN8]: To discuss. It's probably best to say that we will not look at this at all but could say:
"However, exploratory analysis will examine treatment effect heterogeneity for several key baseline variables that are anticipated to be potentially predictive of the treatment effect on the primary outcome to inform the design of the planned full-scale trial."

4.3.2 Mediators

The study is not specifically powered to examine the mechanisms of action of the intervention using mediation analysis and therefore mediation analysis will not be undertaken. Signal for efficacy for putative mediators will be based on the analyses described in section 4.2.2.

Commented [SN9]: To discuss. It's probably best to say that we will not look at this at all but could say:
"However, exploratory analysis will examine the effect of putative mediators using the product of coefficients approach to estimate the indirect effect on Y via putative mediators M1, M2, …. These will be reported with 95%CIs without *p*-values"

4.4 OTHER PLANNED ANALYSES

4.4.1 Adverse events

Self-reported SAEs will be reported descriptively as the total overall number and by type. Deaths will be reportedly separately as the number (%) by treatment group. A rate ratio, with and 95% CI, will be provided to summarise the occurrence of SAEs in the treatment group relative to the control group for the intention-to-treat sample.

4.4.2 Intervention adherence and engagement

For those in the intervention group, the number (%) of participants engaging with the treatment in any way will be reported, as well as the number and (%) of participants engaging with each of the sessions. The median number of sections engaged with along with the interquartile range will be presented as will the number (%) individuals considered to have achieved the minimum effective dose (as defined in the analysis sets section above).

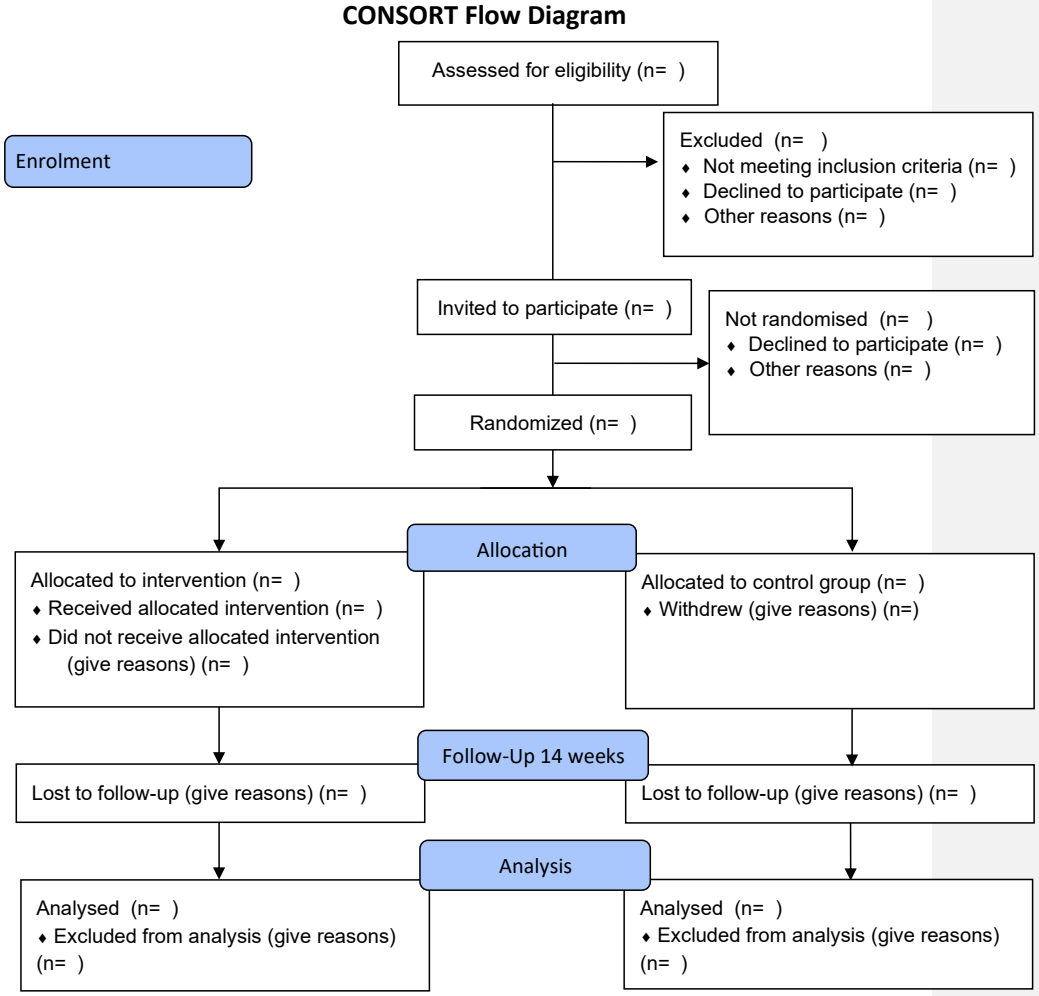
4.4.3 Psychometrics of patient reported outcomes

Commented [SN10]: Question about patient rated validity of the sexual functioning scales used (for total scores)

An objective of the study is to identify appropriate instruments assessing sexual dysfunction as candidates for the primary outcome in a future definitive trial. To inform this the psychometric quality of self-report instruments will be assessed. We will assess reliability using McDonalds Omega, with a minimum accepted cut off ω>0.70 and a preference for values ω>0.80. Reliability calculations will also be computed for sub-scales where relevant (i.e., listed in Section 2.5.3). Potential issues with floor and ceiling effects will be considered, particularly where these might impact responsiveness in the definitive trial. Individual items will be checked for missingness to ensure that there are no problematic items for this population.

5 APPENDIX

5.1 DUMMY CONSORT FLOWCHART



5.2 DUMMY TABLES FOR MAIN PAPER

Tables for the main paper will be provided in full in an MS Excel spreadsheet with one sheet per table. Versions of these tables that will be provided in the main efficacy analysis report in MS Word with minor formatting difference. It is anticipated that the two tables below are included in the main submission to a peer-reviewed journal.

'Table 1': Baseline variables by group

Variable		Group 1 (n=?)	Group 2 (n=?)
Age			
Gender	Man		
	Woman		
	Non-binary/other		
Type	Primary progressive		
	Secondary progressive		
	Relapsing remitting		
Ethnicity	White		
	Black		
	Asian		
	Mixed		
	Other		
..continues..			

'Table 2': Patient reported outcome by group

	Group 1 (n=?)			Group 1 (n=?)			Adjusted mean difference						
	N	Mean	SD	N	Mean	SD	Diff	SE	z	p	95%ll	95%ul	SMD
SFEQ													
MSISQ-19													
FSDS-R													
DSC													
NSSS-S													
GNDS													
PHQ-9													
GAD-7													
FFMQ													
EQ-5D-5L													