The Community Navigator Trial

Statistical Analysis Plan

Version 1.0

31 May 2024

1 Introduction

1.1 Purpose

This statistical analysis plan (SAP) contains details of the main pre-specified statistical analyses for the Community Navigators trial. This SAP describes the statistical analysis of the clinical outcomes; it does not contain details of any qualitative analyses. The SAP does not preclude the undertaking of further ad hoc or exploratory analyses, although the results of any such analyses should be interpreted with caution. Furthermore, the SAP does not preclude the adaptation of any part of the trial analysis should situations arise in which such adaptation is deemed necessary; any such adaptation will be transparent and fully justified.

1.2 Protocol version

Full details of the trial design, population, intervention, comparison and outcome variables may be found in the protocol (version 5.0, 6th December 2023).

1.3 Trial registration

This trial was prospectively registered with ISRTCN (ISRCTN13205972).

1.4 Authorship

This SAP has been written by Gareth Ambler (GA) and Rebecca Jones (RJ), following the guidelines of Gamble *et al* (JAMA 2017. doi:10.1001/jama.2017.18556).

1.5 SAP Revisions

Version	Date	Changes
1.0	31 May 24	

1.6 Signatures

Authorised by:	Signature	Date
Prof. Brynmor Lloyd Evans Chief Investigator	Signed by: Brynmor Lloyd-Evans DCE3BE6A34F74B3...	8/5/2024
Prof. Gareth Ambler Lead Statistician	Signed by: Gareth Ambler 9DB1D61CAC05408...	8/12/2024
Prof. Nicola Wiles DMEC Chair	Signed by: Nicola Wiles 45007E0B42AE4F7...	8/7/2024
Prof. Mike Slade TSC Chair	Signed by: Mike Slade 227B247D3D0B4E3...	8/5/2024

2 Trial Summary

2.1 Title

Randomised controlled trial of the Community Navigator programme to reduce loneliness and depression for adults with treatment resistant depression (TRD) in secondary mental health services (short title: The Community Navigator Trial).

2.2 Aims

The specific objectives of the Community Navigator Trial are to:

- assess the effectiveness of the Community Navigator programme at reducing depression in adults with treatment resistant depression (TRD) by the end of treatment (8 months)
- assess the effectiveness of the Community Navigator programme in relation to depression 6 months after the end of treatment (14 months post-randomisation) and in relation to other outcomes, particularly loneliness, anxiety and personal recovery
- determine the cost effectiveness of the Community Navigator programme
- explore the perceived impact of the Community Navigator programme, how benefits were achieved and key considerations for its provision in NHS settings.

2.3 Population

Adults with treatment resistant depression (TRD).

2.3.1 Inclusion criteria

Adults aged 18 years and over who meet ICD-10 diagnostic criteria for depression, have had at least two reported courses of antidepressants without symptom remission and score 2 or more on the 6 item De Jong Gierveld Loneliness Scale (DJG-6).

2.3.2 Exclusion criteria

Individuals are not eligible to participate if they are due to be discharged from the mental health team within the trial intervention period (8 months), are currently using mental health inpatient services, have a primary diagnosis of schizophrenia, other non-affective psychotic disorder or bipolar disorder, lack capacity to consent, do not understand English well enough to engage with the study intervention or have a care coordinator who supervises the Community Navigators.

3 Study Methods

3.1 Design

A multi-site, single (researcher) blind, parallel group, individually randomised controlled trial to assess the clinical and cost effectiveness of the Community Navigator programme for people with treatment resistant depression using secondary community mental health services.

3.1.1 Intervention

The Community Navigator programme, comprising up to 10 sessions of one-to-one support from a Community Navigator and attendance at up to four participant group "meet ups", plus usual care. The Community Navigator will work with the participant to increase their social activities and community engagement. The Community Navigator programme is provided in addition to 'treatment as usual' (below).

3.1.2 Comparison (treatment as usual)

An information booklet about local social resources, plus usual care.

3.2 Randomisation

Participants will be randomly assigned in a 1:1 ratio to the Community Navigator programme or the control group. Randomisation will be stratified by site using block randomisation with varying block sizes. Full details can be found in the randomisation protocol.

3.3 Sample size

The sample size of 306 participants (153 per study arm) calculated for the trial will allow detection of a 0.4 standard deviation (SD) difference in PHQ-9 depression score between arms with a 2-sided alpha of 5% and 90% power (though see below). The standard calculation to detect a 0.4 SD effect size with 90% power and 5% alpha requires 132 participants per arm. Assuming a correlation with PHQ-9 at baseline of 0.5 and loss to follow up of 15% results in a required sample size of 117 participants per arm. Assuming four sites, three Navigators per site, and an intraclass correlation coefficient (ICC) for clustering by Navigator of 0.05 results in a design effect of 1.6 (based on 13 participants per Navigator). Inflating the sample size for clustering in the intervention arm only (to 188), then adjusting group sizes to be equal, produces a sample size of 153 participants per arm or 306 participants in total.

In response to peer review comments, the sample size calculation was checked using simulation and we found that 90% power would be achieved if the data are analysed using a mixed model estimated using Maximum Likelihood (ML) estimation treating each control subject as a cluster of size 1[1]. However, confidence interval coverage and Type I error are slightly improved by estimating the model using Restricted Maximum Likelihood (REML)

estimation with the Kenward-Rogers adjustment. This results in a slight loss of power but is now the planned analysis. Formally, the power to detect a (standardised) difference of 0.4 is reduced to 87%, or equivalently, we have 90% power to detect a difference of 0.42.

3.4 Internal Pilot and Interim Analyses

3.4.1 Internal Pilot

Trial recruitment and engagement with the Community Navigator programme will be monitored and the continuation criteria will be reviewed during the internal pilot phase of the trial. This process is fully documented in the protocol.

3.4.2 Interim analyses

No interim analyses are planned.

3.4.3 Blinding

This is a single blind trial. Assessors are blind to treatment allocation; participants are not. Statisticians will also be blinded to allocation as far as possible until after the primary analysis has been agreed. One of the statisticians will attend the Data Monitoring and Ethics Committee (DMEC) and consequently may become unblinded if the committee requires any statistics (particularly Adverse Events) to be reported separately by study arm.

3.5 Observation times

Data will be collected at the following time points during the trial:

- baseline
- 4 months post-randomisation
- 8 months post-randomisation (end of treatment and primary endpoint)
- 11 months post-randomisation
- 14 months post-randomisation

Not all measures will be recorded at every time point (see Table 1). The data recorded at baseline and post-intervention time points will constitute the full dataset for the purpose of analysis of the primary outcome.

At each follow up time there is a data collection window. These are +2 months for 4 and 11 month follow ups and +3 months for 8 and 14 month follow ups. Data will be considered recorded at a given time point, provided that these data are collected from each participant within this window. Any participants for whom data are not collected within this time window will be considered missing at that follow up time for the purpose of the primary analysis.

In the event of any data being collected in error outside of the permitted time windows, the number and percentage of observations excluded for being outside the relevant time window will be summarised separately by study arm for the primary outcome only.

4 Statistical Principles

4.1 Organisation of data and analyses

The SAP will be finalised and approved prior to commencing analysis. The programs and code to be used for statistical analyses will be prepared prior to unblinding as far as possible. Two statisticians will perform the primary analysis independently.

Prior to database lock, basic checks will be performed by the statistician on the blinded data to ensure completeness and accuracy. Outcome variables (primary and secondary), baseline demographic variables and key date variables will be checked for:

- missing values
- values outside an acceptable range
- other inconsistencies

If missing values or other inconsistencies are found, the corresponding data will be sent to the Trial Manager for checking and will either be corrected, deemed to be missing or confirmed correct, as appropriate.

4.2 Confidence intervals and p-values

All statistical tests will be two-sided. All estimates will be presented with two-sided 95% confidence intervals.

4.3 Adherence to intervention

Adherence to the intervention is defined as attendance at three or more individual meetings with Navigators (regarded as receipt of the intervention per protocol).

Some participants may see more than one Navigator, which potentially complicates the outcome analyses. Therefore, if a participant has seen multiple Navigators, we will define their Navigator to be the one that they saw the most. If there is a tie, we will define their Navigator to be the one that they saw the earliest (of those Navigators involved in the tie).

4.4 Analysis populations

The ‘intention-to-treat’ population will include all randomised participants according to the treatment to which they were randomised to receive. Any participants that have withdrawn from the trial, and withdrawn permission to keep and use their data, will be necessarily excluded.

5 Trial Population

5.1 Recruitment and retention

A CONSORT diagram will be presented to provide a detailed description of participant numbers at each time point during the trial. In addition, a table summarising the number of participants who have been lost to follow up at each stage of the trial and reasons for loss to follow up (if supplied) will be presented.

5.2 Baseline characteristics

The demographic information collected at baseline will be presented in a table summarised separately by study arm. Categorical variables will be reported as counts and percentages. Continuous variables will be summarised as either means and standard deviations (SD) or medians and interquartile ranges, depending on the distribution of the data. No statistical tests will be performed to assess baseline differences between study arms. In addition, all baseline outcomes (screening, primary and secondary, see Table 1) will be presented in a table summarised separately by study arm.

The following characteristics will be summarised:

- age
- gender
- ethnicity
- marital status
- living arrangements
- employment status and time since last paid employment
- diagnosis
- years since first depressed
- years since first used mental health services
- baseline outcomes (screening, primary and secondary)

Other available data will comprise:

- site identifier
- dates of assessments
- credibility and expectancy
- concomitant mental health treatments and medications
- reasons for withdrawal or loss to follow-up (if supplied)

6 Analysis

6.1 Outcomes

6.1.1 Primary outcome

The primary outcome is depression measured using the Patient Health Questionnaire (PHQ-9) total score at 8 months post-randomisation.

6.1.2 Secondary outcomes

All secondary outcomes will be analysed at all available time-points (typically 8 and 14 months, though see Table 1). These outcomes are:

- depression measured using the PHQ-9 total score at 4, 11 and 14 months post-randomisation.
- We will create two further variables from the PHQ-9 scores for analysis at all time points: a) recovery from depression, where PHQ-9 score will be dichotomised for analysis, where ≥10 is the clinical cut off for depression; and b) substantial improvement in depression: a dichotomous variable for whether or not a reduction in PHQ-9 score since baseline of at least 5 points has been achieved, based on the established threshold for reliable and clinically significant change.
- loneliness measured using the University of California at Los Angeles Loneliness Scale (ULS-8)
- anxiety measured using the Generalised Anxiety Disorder Questionnaire (GAD-7).
- personal recovery measured using the Process of Recovery Questionnaire (QPR)
- multiple identities measured using the Multiple Identity Scale (MIS)
- self esteem measured using the Brief Rosenberg Self-Esteem Scale (B-RES)
- self stigma measured using the Discrimination and Stigma Scale (self-stopping behaviour subscale) (DISC-12)
- social network measured using the Lubben Social Network Scale (LSNS-6)

6.1.3 Safety outcomes

- withdrawal from the intervention and reasons for withdrawal
- serious adverse events checklist

Further details on the scoring and ranges of all outcomes can be found in the Appendix.

6.1.4 Timing of outcomes

Table 1 provides an overview of primary and secondary outcomes and the time points at which they will be collected.

Table 1: Data collection measures and time points (from Table 1 in protocol).

Measure	0	4	8	11	14
Screening measures					
CIS-R depression screening	✓				
DJG-6 Loneliness measure	✓				
Previous anti-depressant use	✓				
Primary outcome					
Depression severity (PHQ-9)	✓	✓	✓	✓	✓
Secondary outcomes					
Loneliness (ULS-8)	✓	✓	✓	✓	✓
Anxiety (GAD-7)	✓		✓		✓
Process of Recovery (QPR)	✓		✓		✓
Four-item Multiple Identities Scale (MIS)	✓		✓		✓
Brief Rosenberg self-esteem scale (B-RSES)	✓		✓		✓
Self-stigma (DISC-12 sub-scale)	✓		✓		✓
Lubben Social Network Schedule (LSNS-6)	✓		✓		✓
Self-reported measure of expectation and credibility of the intervention	✓				
Other measures					
Adverse events and concomitant mental health medication/treatment	✓	✓	✓		

6.2 Primary outcome analysis

The primary analysis of the PHQ-9 score at 8 months follow up (end-of-treatment) comparing intervention and control groups will use a mixed model (estimated using REML, with the Kenward-Roger adjustment) to perform an individual level analysis and will follow guidance in adjusting for Navigator clustering in the intervention arm only (random coefficient model)[1,2]: specifically, each control subject will be treated as a cluster of size 1. This model will also adjust for baseline PHQ-9 score and site using fixed effects. The estimated intervention effect will be reported with a 95% confidence interval and p-value. This analysis will use available data only.

6.2.1 Model checking

All modelling assumptions will be checked. In particular, a confirmatory analysis will be performed using the heteroscedastic model which allows the residual variance for intervention and control groups to differ[1,2].

6.3 Secondary outcome analyses

The effect of the intervention on secondary outcomes will be assessed using analogous methods to those used for the primary outcome. Most of the secondary outcomes are numerical and hence will be analysed using a similar model to that used for the primary outcome, whereas the binary outcomes (those derived from PHQ-9) will be analysed using a mixed-effects logistic regression model. P-values will not be reported for secondary analyses. These analyses will be considered supportive.

6.4 Sensitivity analyses

Several sensitivity analyses may be performed for the primary analysis. These are:

- Analyses may be performed to adjust for any baseline imbalance caused either by chance or by missing data (also see Section 6.6).
- An analysis will be performed that includes PHQ-9 data from all four follow-up time-points. This will use a 3-level mixed model (based on the primary analysis model) with interaction terms between the intervention indicator and time-point. Of primary interest is the estimate at 8 months.

6.5 Subgroup analyses

There are no planned subgroup analyses.

6.6 Missing data

Withdrawals from the study, loss to follow up and other missing outcome data will be summarised separately by randomised group. Potential bias due to missing data will be investigated by comparing the baseline characteristics of participants with and without missing values. Depending on the quantity of missing values, predictors of missingness may be identified. We will then perform a sensitivity analysis that includes these predictors of missingness as covariates in the primary analysis model.

Multiple imputation may also be performed, if deemed appropriate[3]. The imputation model will include outcome data from all time-points, as well as baseline characteristics and any predictors of missingness. The primary analysis model will then be re-fitted using the imputed data.

In addition, imputation may be performed under the assumption that the missing data are missing not at random (MNAR). Two strategies may be investigated:

- Delta-adjustment: This approach initially uses (standard) multiple imputation to impute missing values but then the imputed values are modified using a ‘delta-adjustment’. Different values of delta may be specified, and these can differ by trial arm.

- Reference-based sensitivity analyses: A range of different assumptions regarding the missing data can be investigated using this approach, which is implemented in the Stata package **mimix**[4].

6.7 Additional analyses

The following additional analyses will be performed. These are:

- Adherence to the intervention will be described, e.g. in terms of the mean (SD) numbers of individual and group sessions attended. We will also quantify how many participants changed Navigator.
- A complier average causal effect (CACE) analysis will be performed for the primary outcome to adjust for non-adherence to the intervention. Adherence to the intervention is defined as attendance at three or more individual meetings with Navigators (from Section 4.3).

6.8 Exploratory analyses

Several additional analyses may be performed. These are:

- An analysis will explore whether there is a dose response effect of the number of sessions attended, main results permitting.
- The mediating effect of loneliness on depression will be explored across the five time points, main results permitting.
- The effect of baseline expectations and credibility of the intervention on the outcome will be explored, main results permitting.

6.9 Adverse events

The number, nature and severity of serious adverse events (if any) will be reported separately by study arm at each follow up time point. The number of participants who experience adverse events will likewise be reported separately by study arm.

6.10 Reporting

Analyses will be reported with regard to the CONSORT checklist[5] and with any particular requirements of academic journals and the funders to which the results of analyses are submitted.

7 Appendix

7.1 List of Abbreviations

B-RES	Brief Rosenberg Self Esteem Scale
CI	Confidence interval
CONSORT	Consolidated standards of reporting trials
DISC	Discrimination and Stigma Scale
DMEC	Data Monitoring and Ethics Committee
GAD	Generalised Anxiety Disorder Questionnaire
LSNS	Lubben Social Network Scale
MIS	Multiple Identify Scale
PHQ	Patient Health Questionnaire
SAP	Statistical analysis plan
SD	Standard deviation
TRD	Treatment resistant depression
TSC	Trial Steering Committee
ULS	University of California at Los Angeles (UCLA) Loneliness Scale

7.2 Coding of Outcomes

7.2.1 Depression severity (PHQ-9)

This is calculated by assigning scores of 0, 1, 2, and 3, to the response categories of "Not at all", "Several days", "More than half the days", and "Nearly every day", respectively. PHQ-9 total score for the nine items ranges from 0 to 27. Scores of 5, 10, 15, and 20 represent cut-points for mild, moderate, moderately severe and severe depression, respectively.

- Kroenke et al. (2001)
- https://www.pcpcc.org/sites/default/files/resources/instructions.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495268/

7.2.2 Loneliness (ULS-8)

The ULS-8 has 8 items each scored on a 4-point scale (1=Never, 2=Rarely, 3=Sometimes, 4=Often). These item scores can be summed to give a total score.

- Hays & DiMatteo (1987)
- https://www.researchgate.net/profile/Ronald-Hays/publication/19588637_A_Short-Form_Measure_of_Loneliness/links/58571ced08ae81995eb6b9c6/A-Short-Form-Measure-of-Loneliness.pdf

7.2.3 Anxiety (GAD-7)

This is scored in the same way as PHQ-9. GAD-7 total score for the seven items ranges from 0 to 21. Scores of 5, 10, and 15 represent cut-points for mild, moderate, and severe anxiety, respectively. When screening for anxiety disorders, a recommended cut-point for further evaluation is a score of 10 or greater.

- Spitzer et al. (2006)
- https://www.pcpcc.org/sites/default/files/resources/instructions.pdf
- https://depts.washington.edu/psychres/wordpress/wp-content/uploads/2017/09/gad7_ref.pdf

7.2.4 Process of Recovery (QPR)

The QPR has 15 items each scored on a 4-point scale (0=Disagree strongly, 1=Disagree, 2=Neither agree nor disagree, 3=Agree, 4=Agree strongly). Higher scores are indicative of recovery. These items can be added to give a total score.

- Neil et al. (2009)
- https://www.leedsandyorkpft.nhs.uk/advice-support/wp-content/uploads/sites/3/2019/11/Questionnaire-about-the-Process-of-Recovery.pdf
- https://www.tandfonline.com/doi/full/10.1080/17522430902913450

7.2.5 Four-item Multiple Identities Scale (MIS)

This has 4 items each scored on a 7-point scale (1='Do not agree at all' to 7='Agree completely'). These item scores can be summed to give a total score.

- Haslam et al. (2008)
- https://www.tandfonline.com/doi/pdf/10.1080/09602010701643449

7.2.6 Brief Rosenberg self-esteem scale (B-RSES)

This has 5 items, each scored on a 4-point scale. These are scored as 4=Strongly agree, 3=Agree, 2=Disagree, 1=Strongly Disagree, for items 1-3, and reverse-scored for items 4 and 5. These item scores can be summed to give a total score.

- Monteiro et al. (2021)
- https://link.springer.com/article/10.1007/s11482-021-09936-4

7.2.7 Self-stigma (DISC-12 sub-scale)

The 4-item Stopping Self subscale has 4 items, each scored on a 4-point scale. These are scored as 0=Not at all, 1=A little, 2=Moderately, 3=A lot. The mean score will be calculated. Levels of discrimination will be evaluated against the criteria in Brohan et al. (2013), i.e. DISC mean scores <1: minimal discrimination; 1–1.5: low discrimination; 1.5–2: moderate discrimination; and >2: high discrimination.

- Brohan et al. (2013)
- https://www.sciencedirect.com/science/article/pii/S0165178113001388

7.2.8 Lubben Social Network Schedule (LSNS-6)

This has 6 items, each scored on a 6-point scale (0='0 relatives/friends' to 5='9 or more relatives/friends'). These item scores can be summed to give a total score.

- Lubben et al. (2006)
- https://academic.oup.com/gerontologist/article/46/4/503/623897

8 References

8.1 Methodology

1. Flight L, et al. Recommendations for the analysis of individually randomised controlled trials with clustering in one arm – a case of continuous outcomes. BMC Medical Research Methodology 2016; 16(165).
2. Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. Clinical Trials 2005; 2(2):152-62.
3. Sullivan TR, et al. Should multiple imputation be the method of choice for handling missing data in randomized trials? Stat Methods Med Res 2018; 27(9):2610-26.
4. Cro S, et al. Reference-based sensitivity analysis via multiple imputation for longitudinal trials with protocol deviation. Stata Journal 2016; 16(2):443-63.
5. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340:c332.

8.2 Coding of outcomes

- Kroenke K, Spitzer RL, Williams JB. The PHQ‐9: validity of a brief depression severity measure. Journal of general internal medicine. 2001 Sep;16(9):606-13.
- Hays RD, DiMatteo MR. A short-form measure of loneliness. Journal of personality assessment. 1987 Mar 1;51(1):69-81.
- Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine. 2006 May 22;166(10):1092-7.
- Neil ST, Kilbride M, Pitt L, Nothard S, Welford M, Sellwood W, Morrison AP. The questionnaire about the process of recovery (QPR): a measurement tool developed in collaboration with service users. Psychosis. 2009 Aug 1;1(2):145-55.
- Haslam C, Holme A, Haslam SA, Iyer A, Jetten J, Williams WH. Maintaining group memberships: Social identity continuity predicts well-being after stroke. Neuropsychological rehabilitation. 2008 Nov 1;18(5-6):671-91.
- Monteiro RP, Coelho GL, Hanel PH, de Medeiros ED, da Silva PD. The efficient assessment of self-esteem: proposing the brief rosenberg self-esteem scale. Applied Research in Quality of Life. 2022 Apr;17(2):931-47.
- Brohan E, Clement S, Rose D, Sartorius N, Slade M, Thornicroft G. Development and psychometric evaluation of the Discrimination and Stigma Scale (DISC). Psychiatry research. 2013 Jun 30;208(1):33-40.
- Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, Stuck AE. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. The Gerontologist. 2006 Aug 1;46(4):503-13.

8.3 Missing scores in PHQ-9

- Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. General hospital psychiatry. 2010 Jul 1;32(4):345-59.