Statistical Analysis Plan

Version 2.0

Trial Title A randomised controlled trial of the clinical and cost-effectiveness of the Journeying through Dementia intervention compared to usual care

Short Title Journeying Through Dementia (JtD) Trial

Funder and Reference NIHR HTA 14/140/80

Sponsor Sheffield Health and Social Care NHS Foundation Trust

Authored by

_____ ___/___/_____ *Ellen Lee Date*
Trial Statistician
Sheffield CTRU, University of Sheffield

Approvers

_____ ___/___/_____ *Prof Stephen Walters*
Date
Senior Trial Statistician
ScHARR, University of Sheffield

_____ ___/___/_____ *Prof Gail Mountain Date*
Chief Investigator
University of Bradford

_____ ___/___/_____ *Prof Catherine Hewitt*
Date
TSC Chair and Senior Statistician
University of York

_____ ___/___/_____ *Emily Robinson Date*
DMEC Statistician
Kings College London

Table of Contents

SAP amendments since version 1

Version	Date approved	Modifications	Prior to/after blind review	Prior to/after unblind review
2.0		- Changed to reflect SAEs not collected for participating supporters (section 9.13) - Changed to reflect DEMQOL and EQ-5D-5L collected either by phone or face to face (sections 11.2, 11.10) - Removed smoking status from example table as not collected (section 14)	Blind review has not been conducted, however blinded data has been see in order to start developing programs	Prior to unblind review

List of Abbreviations

AE	Adverse Event
CACE	Complier Average Causal Effects Analysis
CI	Confidence Interval
CONSORT	CONsolidated Standards Of Reporting Trials
CRF	Case Report Form
CTRU	Clinical Trials Research Unit
DMP	Data Management Plan
eCRF	Electronic Case Report Form
DEMQOL	Dementia related quality of life measure
DMEC	Data Monitoring and Ethics Committee
EQ-5D-5L	Measure of Health Status
GAD-7	Measure of Generalised Anxiety Disorder
GLM	Generalised Linear Model
GSE	General Self-Efficacy Scale

HTA	Heath Technology Assessment
HRQoL	Health related quality of life
IADL	Instrumental Activities of Daily Living
ICC	Intraclass correlation coefficient
ICH	International Conference of Harmonisation
IQR	Interquartile range
ITT	Intention-to-treat
JtD	Journeying through Dementia
LTFU	Lost to follow-up
MI	Multiple Imputation
MMSE	Mini Mental State Examination Score
NIHR	National Institute for Health Research
NHS	National Health Service
OR	Odds ratio
PHQ-9	Patient Health Questionnaire
PP	Per-protocol
PROM	Patient reported Outcome Measure
QoL	Quality of Life
RCT	Randomised Controlled Trial
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCQ	Sense of Competence Questionnaire
SD	standard deviation
SMAS	Self-management ability scale
TAU	Treatment As Usual
TMG	Trial Management Group

TSC Trial Steering Committee

List of Tables

List of Figures

Definition of Terms

Participant- This refers to a person with dementia who is participating in the trial.

Participating supporter- This is a family member, friend or neighbour that provides support to a person with

dementia. They may be known as a 'carer'. In the trial, participating supporters are people that have consented into the trial to complete outcome measures. They may also help a person with dementia participate in the trial, such as liaising with researchers to organise visits; and participating in the intervention if allocated to receive it.

Supporter- This is a family member, friend or neighbour that provides support to a person with dementia. They may be known as a 'carer'. In the trial, supporters are people that may be helping a person with dementia participate in the trial, such as liaising with researchers to organise visits or attending the intervention if allocated to receive it. However supporters are not participating in the trial themselves, for example no outcome data is collected from them.

1. Introduction and Trial Objectives and Design

This Statistical Analysis Plan (SAP) provides detailed guidance of the statistical analysis for the Journeying through Dementia (JtD) trial. This section provides a brief background of the trial, the primary research question under investigation, study design used to address the research questions, and key documents guiding the development of this SAP.

1.1 Brief Trial Background and Primary Research Question

There is an increasing burden of dementia in the UK and globally impacting those living with the condition, and their family carers, services and the economy. Whilst existing research has provided insights into the potential benefits of promoting self-management for people with dementia, there is an absence of robust evidence through full-scale RCTs. This means that it is difficult to establish the effectiveness and cost-effectiveness of such interventions, particularly in comparison to usual care.

The JtD intervention has been designed to support people in the early stages of dementia and improve their quality of life (QoL) by promoting self-efficacy and assisting them to continue to participate in life and maintain their independence. The JtD involves individuals in the early stages of dementia participating in 12 facilitated weekly group sessions and in 4 individual sessions with one of the facilitators. It is expected that the timing of the individual sessions will be one before, one after and two during the course of the group sessions. The group is encouraged to select the content of their sessions from a range of topics including strategies to manage memory challenges, engaging in hobbies/interests and ways of maintaining physical and mental wellbeing. An essential component is the enactment of activities in the community with support from each other. During individual sessions people with dementia are assisted to work on individual needs and goals. Participants are not necessarily required to nominate a supporter (family member or friend who provides them with support) to take part, but if supporters are involved they

are invited to join group sessions one, six and twelve and can participate in the individual sessions with the person with dementia if agreed.

This trial therefore aims to determine the clinical and cost-effectiveness of the JtD intervention for people in the early stages of dementia. The trial is funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme, and the Sheffield Health and Social Care NHS Foundation Trust acts as the trial sponsor.

1.2 Trial Design

This study is a pragmatic, two-arm, parallel group, superiority, individually and equally randomised controlled trial (RCT) comparing the JtD intervention with usual care to determine benefit for people in the early stages of dementia. Although the study is an individually RCT, the delivery of the JtD intervention in one arm of the trial is group based. That is, groups of 8 to 12 participants randomised to the JtD intervention will attend 12 weekly facilitated meetings in local venues in order to receive the intervention.

The trial has been designed with an internal pilot phase during the first 8 months of active recruitment which will assess the feasibility aspects of conducting the main trial as guided by pre-planned STOP/GO criteria. For consistency throughout this SAP, the active intervention group shall be referred to as “JtD” and the control group as “Usual Care”.

1.3 Documents Guiding the SAP

This SAP is written in conjunction with the International Conference of Harmonisation (ICH) topic E9 on Statistical Principles for Clinical Trials (ICH, 1998), applicable Standard Operating Procedures from the Sheffield Clinical Trials Research Unit (CTRU), (particularly ST001), and the Trial Protocol and related amendments.

1.4 Main Aims and Trial Objectives

As highlighted in Section 1.1, the primary aim of the trial is to determine the clinical and cost effectiveness of the JtD intervention compared to usual care for people in early stages of dementia. The specific objectives are to:

a) conduct an internal pilot RCT to assess the feasibility of rates of recruitment considered,

b) proceed to the full trial if the internal pilot trial feasibility criteria are met,

c) conduct fidelity checks regarding the delivery of the JtD intervention,

d) conduct an embedded qualitative nested study to explore issues concerned with intervention delivery,

e) identify how the intervention might be realistically delivered through services.

2. Scope of the SAP

This trial has been designed with an internal pilot phase to assess the feasibility of patient recruitment. Furthermore, there is a fidelity study to assess the delivery of the JtD intervention and an embedded qualitative sub study to explore issues concerned with the delivery of the intervention. The trial is also designed with health economic evaluation to address the cost effectiveness of the JtD intervention. This SAP, however, focuses only on addressing the internal pilot objectives and clinical effectiveness research questions of the trial. Therefore, the fidelity and embedded qualitative studies and health economic evaluation aspects which are out of scope of this SAP will be addressed elsewhere.

3. Outcomes Measures and Timing

This section describes the outcome measures, which are used to evaluate the feasibility, primary, and secondary objectives of the trial relating to the clinical effectiveness of the JtD intervention compared to usual care. The timing of all outcome measures is explicitly stated. Outcomes relating to fidelity and embedded qualitative studies are excluded, as they are out of scope of this SAP. Although health economic evaluation is out of scope of this SAP, related outcomes are stated here for completeness as their analysis will be reported elsewhere.

3.1 The Internal Pilot Primary Outcomes

The following feasibility outcomes will be reviewed after 8 months of active recruitment to assess the STOP/GO criteria:

- Recruitment of a minimum of 113 participants across the six pilot sites by the end of the fifth month of active recruitment (75% of the 150 target).

- Recruitment of a minimum of 12 facilitators (two facilitators identified at each of the six pilot sites by the start of active recruitment to deliver the intervention).

- No more than two of the six planned groups in the internal pilot with less than four participants registered for the group by the sixth month of active recruitment.

3.2 Primary Outcome

To address the primary clinical effectiveness research question of the JtD intervention, the primary endpoint assesses the health related QoL in participants measured by the DEMQOL (Smith et al, 2005) which is evaluated at 8 months post randomisation. The scoring system to generate summary scores for analysis is detailed in Section 11.

3.3 Secondary Outcomes for Participants

Seven secondary PROMs are assessed at 8 months post-randomisation focusing on participants. Scoring systems used to compute summary measures and dealing with missing data are detailed in Section 11: a) PHQ-9; measure depression severity (Hancock et al, 2009)

b) Generalised Anxiety Disorder Assessment (GAD-7); severity measure of generalised anxiety disorder (Wild et al, 2014)

c) Instrumental Activities of Daily Living (IADL); measures independent living skills (Gold, 2012) d) General Self Efficacy (GSE); measures general perceived self-efficacy (Schwarzer et al, 1995) e) Diener's Flourishing Scale; measures psychological flourishing (Diener et al, 2010)

f) Self-Management Ability Scale (SMAS); measures self-management abilities (SMA) (Schuurmans, 2005)

g) EQ-5D-5L; measures health outcome (Brazier et al, 2007)

Additional secondary PROMs and resource use outcomes are assessed at 12 months post-randomisation to further explore the clinical and cost-effectiveness research questions.

a) DEMQOL

b) EQ-5D-5L

c) Health and social care resource use questionnaire (for use in health economic evaluation)

3.4 Secondary Outcomes for Participating Supporter

The following secondary outcomes will be assessed at 8 months post randomisation focusing on the participating supporter:

a) PHQ-9

b) Sense of Competence Questionnaire (SCQ); measures sense of competence in caregivers (Vernooij Dassen, 2003)

c) EQ-5D-5L

3.5 Safety and Harms Outcomes

There are few anticipated adverse effects of the JtD intervention. Adverse Events (AEs) are not anticipated as a consequence of the intervention and thus will not be recorded. Serious Adverse Events (SAEs) will be recorded for all participants. Six categories of SAEs will be recorded during follow up:

(a) results in death

(b) is life-threatening (subject at immediate risk of death)

(c) requires hospitalisation or prolongation of existing hospitalisation

(d) results in persistent or significant disability or incapacity

(e) consists of a congenital anomaly or birth defect

(f) is otherwise considered medically significant by the investigator

4. Sample Size Estimation

As stated in Section 3.2, the primary outcome for the study is the mean DEMQOL score 8 months post

randomisation. If we assume a standard deviation of 11 points for the DEMQOL, and that a mean difference of 4 or more points is clinically and practically important (Mulhern, 2013). The sample size has been calculated to have a 90% power of detecting this 4-point difference or more, if it truly exists, which is equivalent to a standardised effect size of 0.36 in group mean scores at 8 months, as being statistically significant at 5% two sided level. As the JtD intervention is a facilitator led group intervention, the success of the intervention may depend on the facilitator delivering it so the outcomes of the participants in the same group with the same facilitators may be clustered. For an individually RCT without adjustment for clustering by facilitator, the target sample size would be 160 per arm (a total of 320). Assuming an average cluster size of 8 dementia patients per facilitated group (Mountain, 2017) and an intra-cluster correlation (ICC) of 0.03, this will inflate the sample size by a design effect of 1.21; to 194 per group (a total of 388) with valid primary outcome data. We further assumed 20% lost to follow-up. Therefore, given these calculations, the trial target sample size is to randomise to 243 participants in each arm (a total of 486).

The Dementia Quality of Life measure (DEMQOL) (Mulhern et al 2013) is a 28 item self-completed quality of life questionnaire for people with dementia. The 28 items are scored on a 1 to 4 Likert scale (a lot; quite a bit; a little; not at all) and the total score ranges from 28 to 112. There is no reported or established minimum clinically important difference (MCID) for the DEMQoL so it is difficult to specify a target difference that is clinically and practically important.

Data from the development and validation of the preference/utility based DEMQOL-U outcome (Mulhern et al 2013)) which compared the performance of the DEMQOL (and DEMQOL-U) with other patient reported outcome measures for people with dementia; for example MMSE (Mini Mental State Examination) (Folstein et al 1975); BADLS (Bristol Activity of Daily Living Scale) (Bucks etal 1996), NPI (Neuropsychiatric Inventory) (Cummings et al 1994) with established MCIDs reported changes on the DEMQOL of 5.4; 7.5 and 6.4 points respectively for patients who improved by more than the MCID on the anchor measure. This suggests that our proposed 4-point difference, although small is likely to be of clinical and practical importance.

5. Trial Features to Minimise Bias

This section describes design measures put in place to avoid the potential of bias in assessing the effectiveness of the JtD intervention, focusing on randomisation, its concealment, and blinding of outcome assessors and other research staff. Additional measures adopted to minimise bias during statistical analyses, such as dealing

with missing data, compliance to the JtD intervention, clustering in one arm and stratification during randomisation are addressed in Section 10.

5.1 Randomisation, Allocation Concealment and Blinding of Outcome Assessments

Participants are randomised 1:1 to receive either JtD intervention or Usual Care. Randomisation stratified by stratification site (see section 11 for a definition of stratification site) and blocked by a fixed block size will be undertaken using a web-based randomisation system. The objective is to minimise allocation predictability and imbalance in the distribution of participants between intervention arms and also within delivery site. A fixed block size was used to ensure the imbalance between treatment groups was controlled and minimal; at the start of the study only two JtD intervention groups per site were deemed feasible, and a maximum number of 13 participants allowed in each JtD intervention group. The risk of predicting the allocation ratio in this study is minimal as randomisation is done centrally by Sheffield CTRU staff members and research assistants responsible for recruiting participants at site are blinded to treatment allocation.

Due to the group nature of the intervention, a process of delayed randomisation will be employed whereby participants are not randomised at the point of consent but after the collection of baseline measures (less than 2 months before the intervention wave at that site begins). This will enable operational feasibility in the organisation of the JtD group sessions to minimise delay in intervention delivery. It also means that the time between randomisation and starting treatment is roughly similar (within 8weeks) across participants, and that treatment occurs close to the beginning of the

baseline to 8 month follow up window. In the event of a couple in the same household both consenting to take part in the study the pair will be randomised as a couple and not separately i.e. to both get the intervention or to both get usual care.

A number of measures have been undertaken to minimise potential bias. Randomisation and its concealment will be done using a centralised web-based randomisation system. Furthermore, since the trial uses a PROM as a primary outcome, it is imperative to maintain the blinding of outcome assessors as much as possible during the course of the trial. In this regard, randomisation will be performed centrally by staff at the Sheffield CTRU rather than site staff. Furthermore, to ensure the outcome assessors are blinded to group allocation, the trial support officer, clinical research assistant, trial manager or other member of the study team who will not be conducting outcome assessment will enter the participants' details onto a remote web-based randomisation system. Trial Statisticians and

Health Economists will be blinded during the course of the trial until data freeze. Any 'unblinding' cases for various circumstances will be reported as described in Section 9.7 for transparency and to enhance the interpretation of results.

The reports to the Data Monitoring and Ethics Committee (DMEC) will be prepared by Data Management who are unblind. Unblinded statistical reports may also be provided to the DMEC on their request, as guided by the DMEC Charter by a Sheffield CTRU Statistician who is external to the trial.

6. Data Monitoring and Interim Analyses

The conduct of this trial will be guided by three committees as governed by Sheffield CTRU SOPs, trial protocol, and the DMEC Charter. The committees are:

a. Trial Management Group (TMG)

b. Trial Steering Committee (TSC)

c. DMEC

This trial has been designed as a fixed sample size design with one formal statistical analysis at the planned scheduled end. Therefore, there are no planned interim analyses to allow early stopping using formal statistical rules. However, the trial will be independently monitored by the DMEC within the premise of the related Charter agreed and signed by all the members. A recommendation to stop the trial could be made by the DMEC based on safety reasons as stipulated in the DMEC Charter. Although unlikely to happen, on DMEC's discretion, they may request futility analysis using stochastic curtailment but there is no option for efficacy early stopping.

7. Data Sources and Data Management

Data used in this study will come from data entered onto Case Report Forms (CRFs) and questionnaires. The data are stored on the Sheffield CTRU database system, with the exception of the randomisation list which is held on the CTRU's randomisation system. The CTRU database system has restricted access to certain trial staff depending on their trial related duties and responsibilities. The data management unit of the Sheffield CTRU will validate and query electronic data for inconsistencies during the course of the trial as governed by the processes and procedures stipulated in the Central Data Validation SOP (DM005). The Trial Statistician will conduct any additional

validation checks where appropriate before the data lock and sign off guided by the relevant SOPs (DM005, DM012).

8. Definition of Analysis Populations

This section states and defines the primary analysis populations and other secondary analysis populations which will be used mainly for sensitivity analysis.

8.1 Analysis populations for participants

To minimise potential selection bias and provide unbiased estimate of the treatment effect and to preserve randomisation; primary analysis population will be based on Intention-to-treat (ITT) principle as defined in Table 1. In order to truly implement ITT analysis, missing primary outcome data will be imputed and included in a sensitivity analysis, as described in Section 9.8.2.

Table 1: Definitions of Analysis Populations for participants.

Analysis Population	**Patient inclusion criteria**
Intention-to-treat (ITT)	(1) All participants randomised to receive either JtD intervention or Usual Care **and**, (2) Consented to take part in the study **and**, (3) Treatment assignment during analyses is as allocated at randomisation regardless of what happens after randomisation (such as protocol deviations and withdrawals) **and**, (4) With complete DEMQOL data at 8 months post randomisation (see Section 11.2 for a definition of a complete score). Excluding participants who withdraw before randomisation, and including participants found to be ineligible post randomisation (Altman, 1991).
CACE	a) A subset of the ITT analysis population b) Subgroup of participants who are believed to comply with the JtD intervention as described in Section 9.6 and 10.2, c) Excludes ineligible participants randomised by error d) Includes participants who were randomised to usual care but received and complied with JtD intervention

8.2 Analysis population for participating supporters

This includes participating supporters who met eligibility criteria as stated in the protocol and provided informed consent. Since this is a supplementary analysis, only participating supporters with available follow-up data will be analysed; thus, no imputation will be employed.

9. Outline of the Statistical Analysis

This section outlines the statistical analysis framework to be adopted beginning with how the trial data and results will be reported. The description of the statistical methods used to analyse outcomes to address the research questions are given in order of importance, starting with the primary outcome followed by secondary outcomes, including safety. Dummy tables can be found in section 14 and are provided as an example and to guide the Trial Statistician(s) during analysis and reporting.

9.1 Reporting Framework of Trial Data

The analysis of trial data and reporting will be guided by the revised CONSORT statement for individually randomised parallel group trials (Moher et al. 2010; Zwarenstein et al. 2008) and the extension to the consort statement for non-pharmacologic treatments (Boutron et al, 2008).

It should be highlighted that the trial is also partially clustered in one arm. Thus summaries relating to this clustering within the JtD arm will be reported; for example, number of JtD courses and their attendance, and the observed ICC (with 95% CI) to help the planning of future related partially clustered trials.

9.2 Internal Pilot Analysis

The Trial Steering Committee (TSC), tasked to 'independently' oversee conduct of the trial on behalf of the funder and sponsor, will assess the feasibility outcomes at the end of the pilot phase. They will assess whether the trial should

continue in light of the feasibility results against the STOP/GO criteria. The results and recommendations will be communicated to the funder (NIHR HTA).

9.3 The CONSORT Flowchart: Data Completeness and Disposition

The reporting of data completeness is an integral part of good practice in trial reporting. Guided by the

CONSORT statement for individually randomised parallel group RCTs, a flowchart showing participants (not the participating supporters) from screening, during follow-up, and to inclusion in the primary analysis will be constructed as illustrated in Figure 1. This will also be made available to the trial monitoring committees during the trial. Summaries may be presented by centre and/or overall (and treatment group where appropriate). Data completeness will be based on the primary outcome, DEMQOL score.

9.4 Demographics and Baseline Characteristics of Participants

Summaries of the baseline variables relating to socio-demographics, medical history and characteristics of participants as captured on the CRFs will be presented by intervention group and overall (as in Table 4, Table 5, Table 6, Table 7). The list of baseline variables to be presented is given in Table 2.

No statistical significance testing will be done to test baseline imbalances between the intervention arms but any noted differences will be descriptively reported. Baseline imbalances will be considered for adjusted analysis as part of the sensitivity analysis described in Section 9.8; (the TMG will discuss baseline imbalances, and deemed potentially clinically relevant by the chief investigator will be included in the sensitivity analysis).

Summaries will be presented depending on the distribution of the variable under consideration. For example, minimum, maximum, mean and standard deviation (SD) or median, Interquartile Range (IQR), minimum and maximum will be presented for continuous variables depending on the skewness of the underlying data. Categorical variables will be summarised using numbers and related percentages in each category by intervention group and overall. As for count variables, a decision on the presentation approach will be made based on the underlying distribution of the pooled data. For instance, if the maximum number of counts is small, then a categorical variable will be derived and presented accordingly. Otherwise, the median (IQR) of the distribution of the count variable will be presented.

Table 2: Baseline Data to be presented

Participant Socio-demographics	Categorical variables • Delivery site • Sex • Ethnicity • Living with others (yes/no) • Lives with (spouse, children, etc)

	• Accommodation type Continuous variables • Age (years)

Participant medical history	Categorical variables • Specific health problem • Type of dementia diagnosis Continuous variables • Length of time since dementia diagnosis (years)
Participating Supporter baseline characteristics	Categorical variables • Sex • Relationship to person with dementia • Length of time caring (categories) • Living with person with dementia
Participant Baseline quality of life	Continuous outcomes • DEMQOL (total score and Q29) • EQ-5D-5L (value index and VAS) • PHQ-9 • GAD-7 • GSE • Diener's Flourishing Scale • SMAS • IALD
Participating Supporter baseline quality of life	Continuous outcomes • PHQ-9 • EQ-5D-5L (value index and VAS) • SCQ

9.5 Characteristics of completers and non-completers

The objective of this section is to explore the pattern of missing data of participants and whether completers are systematically different from non-completers. Completers are participants meeting the ITT criteria, who have complete primary endpoint data (DEMQOL at 8 months from randomisation).

Descriptive statistics of baseline variables will be presented stratified by the intervention group and missing data status as illustrated in Table 9 and Table 10. The baseline variables include age, sex, smoking status, type of dementia and QoL measures. Other variables presented in baseline tables may be considered at the discretion of the Trial Statistician. The relationship between treatment compliance and missing data status will be tabulated for the participants in the Intervention group. Predictors of missing data will be included in multiple imputation equations.

9.6 Compliance to JtD Intervention

This section aims to summarise the attendance of participants to group and individual sessions which are part of the JtD intervention. In addition, details about the accompanying persons and their relationship with the participants

will also be summarised. Participants in the JtD arm are expected to attend 12 group sessions and 4 individual sessions; a total of 16 possible sessions. Participants are able to invite a supporter to participate in group sessions 1, 6 and 12, and in the individual sessions if the participant finds this helpful in achieving their goals. As stated in the protocol, a person with dementia will be classified as having received the JtD intervention if they attended at least 10 of the possible 16 sessions (62.5%). The list of JtD intervention compliance summaries is presented in Table 3.

Table 3: Descriptive JtD intervention compliance summaries

General summary (N)	• Total number of JtD groups
Continuous (mean, SD, median, IQR, min, max)	• Session size • Number of sessions attended (group + individual) • Number of group sessions attended • Number of individual sessions attended • Proportion of group sessions accompanied
Categorical (n, %)	• Reason for non attendance • Reason for intervention withdrawal • Number of group sessions accompanied (0,1,2,3) • Number of individual sessions accompanied (0,1,2,3,4) • Accompanying person relationship to participant (participating supporter, non-participating supporter, consultee or other)

Data visualisation approaches will be used to present data in appropriate figures or graphs. Graphical methods and correlation will be used to investigate a relationship between the number of JtD sessions attended and DEMQOL score at 8 months.

Exploratory descriptive analysis will compare those that 'received' the intervention to those that did not (within the intervention arm) with respect to baseline data (age, type of dementia etc.) and socio-economic status as defined by the ONS Index of Multiple Deprivation calculated using participant's postcode data.

9.7 Unblinding of Treatment Allocation

As highlighted in Section 5.1, outcome assessors will be blinded to intervention allocation in order to minimise operational bias. For some reasons, unblinding of treatment allocation may happen. It is therefore important to report the following summaries by intervention group:

a) Number and proportion of unblinded cases,

b) Who was unblinded (outcome assessors, Trial Statisticians, TSC, Health Economist, etc),

c) Method of unblinding (face to face, email, by observation, etc),

d) Source of unblinding (facilitator, participant, DMEC, etc) ,

e) Reason for unblinding,

f) Suspected or real (see section 11.1).

9.8 Efficacy analysis of the intervention

This section describes the statistical methods to be used for the primary analysis and related sensitivity analyses to address the primary research question. The Usual Care arm will be the reference group for the analysis unless stated otherwise.

9.8.1 Primary Endpoint Analysis

The primary aim of the trial is to determine the benefit of an occupational therapy based self-management intervention (JtD) compared to Usual Care for people in the early stages of dementia. To address this aim, the primary analysis will compare mean patient reported DEMQOL total scores at 8 months post randomisation between the JtD intervention and Usual Care groups using a mixed effects linear regression model adjusted for DEMQOL baseline total score and stratification site, and allowing for the clustering of the outcome by the JtD intervention group (Baldwin et al., 2011; Roberts et al., 2015; Walters, 2010). The trial design is partially nested where there is clustering in one arm (JtD) only. Each participant in the unclustered Usual Care arm will be treated as a cluster of size one (a singleton). The cluster indicator will be treated as a random effect and stratification site (a stratification and centre effect variable used for randomisation) as a fixed factor. These two factors will be included in all statistical models.

A partially clustered mixed effects linear regression model with homoscedastic errors as well as a heteroscedasticity mixed effects linear regression model will also be considered to account for potential differential variability of outcomes between the two intervention groups. The final model will be selected based on whether the assumption of homoscedastic errors is reasonable. Degrees of freedom will be computed using Satterthwaite approximation to account for a potential of a small to moderate number of clusters (Satterthwaite, 1946). A 95% confidence interval (CI) for the mean difference in DEMQOL total scores between the JtD and Usual Care arm will be reported together with the associated P-value.

A further adjusted analysis may also be performed alongside this baseline DEMQOL and site adjusted analysis depending on the observed degree of imbalance of important baseline covariates which are or may be of

potential prognostic importance using a mixed effects linear regression model as described for the primary analysis. Additional baseline covariates of prognostic importance will include age, sex, PHQ-9 (total score), and GAD-7 (total score).

9.8.2 The inclusion of couples into the study

It is possible that two people living with dementia in a couple within the same household will want to take part in the study. These participants will be randomised as a couple (i.e. both to intervention or both to control). In the event that there are more than 10 couples (20 participants) the primary and secondary analyses will be changed to take into account the hierarchical or clustered nature of the data. A multi level mixed effects model will be used; the random effects will be JtD intervention groups (top level) and couple/singles (lower level). Individual participants who are not part of a couple will be treated as clusters of size one. We found in a previous similar study (18 couples recruited, total sample size 288) that there was no difference in the primary effectiveness analysis allowing for clustering by couples or not (Mountain, 2017).

If there are 10 couples or fewer, the primary analysis will be as is stated in section 9.8.1; a random effects model with one level of clustering (the JtD intervention groups). This means the clustering of outcomes within couples will be ignored. A sensistivity analysis will be performed using the multi level hierarchical model that includes couple/singles (as clusters of size 2 or 1) will be performed to check the similarity of results.

9.8.3 Sensitivity Analyses for the Primary Endpoint

Sensitivity analyses on the primary outcome will be undertaken based on imputation, CACE and mistimed measurements (MM) populations. The CACE population has been described in Section 8.1. Additional imputation and MM populations are defined as follows:

- Imputation: All participants who were consented and randomised (ITT) including those with missing primary outcome data, excluding those that died before 8 month follow up. For participants that do not have complete primary outcome data, their DEMQOL 8 month score will be imputed by regression imputation and multiple imputation (MI). See section 10.4 for details of these imputation methods.
- Mistimed Measurements (MM); a subset of the ITT population which excludes DEMQOL measurements which were collected outside the acceptable window. Ideally, outcomes will be collected < 2 weeks

pre and <8 weeks post-date they are due. Outcomes collected outside this window will be excluded for this sensitivity analysis.

The primary analysis will be repeated for these data sets and displayed alongside the intention to treat analysis results. Details of CACE analysis are provided in Section 10.2. Sensitivity analyses results will be reported as shown in Table 12 (including the primary ITT set) and presented using a forest plot (Cuzick, 2005), as illustrated in Figure 2.

9.9 Secondary Efficacy Analyses: Participants

Secondary outcomes including the EQ-5D-5L, PHQ-9, GAD7, GSE, Diener's Flourishing Scale, SMA and IADL at 8 months post-randomisation will be compared between the intervention and control groups using a mixed effects linear regression model as for the primary outcome. A 95% CI for the mean difference in this parameter between the treatment groups will also be calculated together with the associated P-value, and 95% CI. A similar approach will be used to compare secondary outcomes at 12 months post randomisation.

9.10 Secondary Efficacy Analyses: Participating Supporter

Outcome measures for the participating supporters will be collected at baseline and 8 months. This includes the PHQ-9, EQ-5D-5L and SCQ. The 8 month outcomes will be compared between the intervention and control groups using a least squares linear regression model. Participating supporters are invited to attend up to 3 sessions, correlation between participating supporters within the same cluster is likely to be minimal, hence a simpler model is preferred and no adjustment for JtD cluster group will be made. The mean difference in outcome with associated 95% CI and P-value will be presented for: a) analysis adjusting for the baseline outcome (i.e. PHQ-9) and stratification site and b) adjusted analysis with additional covariates (in addition to a)) if baseline imbalances have been observed.

9.11 Subgroup analysis

The main objective of this section is to explore heterogeneity in the intervention effects in pre-specified subgroups. An exploratory subgroup analysis will be performed using mixed effects linear regression with the primary outcome, the mean DEMQOL score at 8-month post-randomisation, as the response. An interaction statistical test between the randomised intervention group and subgroup will be used to directly examine the strength of evidence for

the difference between treatment group (JtD versus Usual care) varying between subgroups. Two subgroups of interests for exploratory analysis have been pre-specified based on clinical advice:

- Type of dementia
- Presence of participating supporter (yes/no, see section 11.1).

Subgroup analysis will be performed regardless of the statistical significance on the overall intervention effect. The mean difference in DEMQOL (95% CI) will be computed for each subgroup category and visually displayed using a forest plot (Cuzick, 2005). The regression coefficient for the interaction between treatment group and subgroup will be

presented with the associated confidence interval and P-value. We will not calculate separate p-values within each subgroup category (Assmann et al., 2000; Pocock et al., 2002; Wang et al., 2007). Results will also be presented as shown in Table 8. No adjustment for multiple testing will be made to the type 1 error rate.

The categorisation into subgroups for dementia type is not straightforward; dementia diagnosis depends on the methods used at each site and we anticipate that a large amount of participants will have Alzheimer's, which may leave the other subgroups small. Four options are recorded in the CRF (Alzheimer's, Vascular dementia, Mixed Alzheimer's/vascular dementia, other (with free text details)). Subject to adequate group sizes, three categories of dementia will be used for the formal subgroup analysis (Alzheimer's VS any vascular dementia (either vascular dementia or mixed dementia) VS other). A sensitivity analysis using different clinically relevant group combinations will be performed (using the groups;

Alzheimer's VS vascular dementia VS mixed + other, and; Alzheimer's + mixed VS vascular VS other). Descriptive statistics of Mean DEMQOL per group and treatment difference with 95% confidence interval will be presented for each dementia type presented including smaller groups created based on coding the 'other' category (such as dementia with Lewy bodies).

9.12 Assessment of follow up time in relation to outcome

A process of delayed baseline and randomisation will be implemented on this study to ensure that the delay between baseline measures and a JtD group starting is within 2 months. This means the time between conducting baseline measures and a course starting will vary across participants. There are also possible delays between baseline and randomisation as randomisation is performed centrally at Sheffield. The time between baseline,

randomisation and starting treatment (for JtD arm) will be summarised by arm. Two sensitivity analyses will be conducted on the primary outcome:

- A covariate of days between baseline and group starting (which will be 0 for the control group) will be included in the analysis model.
- A covariate of days between baseline and 8 month post randomisation follow up visit will be included in the analysis model.

9.13 Analysis of Safety and Harms

The following descriptive statistics of serious adverse events (SAEs) will be calculated and reported by treatment group and overall. SAEs will be reported on an intention to treat basis (i.e. according to the group to which the participant was randomised) for participants only. No formal statistical tests will be performed on SAE data. All detailed listings of SAEs

will be made available to the DMEC for each meeting and on their request at any point during the trial. All SAEs will be reported on an intention to treat bases (i.e. according to the group to which the participant was randomised). The Following summaries will be presented (overall and by group)

- Number (%) participants experiencing ≥1 SAE
- Number of all SAEs (including repeated events)
- Seriousness (Death, Life threatening, Inpatient hospitalisation, Prolongs hospitalisation, Persistent or significant disability/incapacity, congenital abnormality/birth defect)
- Intensity (mild, moderate, severe)
- Relationship to study intervention (definite, probable, possible, unlikely, unrelated, not assessable)

In order to further investigate potential harms, the proportion of participants and participating supporters considered to be suffering from moderately severe depression symptoms (based on a PHQ-9 score of 15 or above; Kroenke, 2001) will be presented by treatment group and overall for baseline, 8 and 12 months. The proportion of participants considered to be suffering from moderate/severe levels of anxiety (based on GAD-7 score of 10 or above; Spitzer, 2006) will be presented in the same manner.

10. Detailed Statistical Methods and Calculations

Due to the nature of the study design, there are a number of statistical issues that should be addressed or accounted for during statistical analyses. This section aims to highlight these issues and recommend a framework to address them.

10.1 Stratification Factors and Multicentre Data

The stratification site within recruiting trusts is the only stratification variables. Few recruiting trusts are expected to have more than one stratification site. Adjusting for stratification site during the analysis will account for stratification and also potential centre effect while preserving degrees of freedoms.

10.2 Complier Average Causal Effect Inference

One research question of interest is on the nature of the relationship between adherence to the JtD intervention and benefits of dementia self-management as assessed by health related QoL (DEMQOL).

CACE analysis is an attempt to compare the protocol compliers in the JtD group (those that attended at least 10 out of the possible 16 sessions, as detailed in Section 9.6) to those in the usual care group who are 'likely' to have complied had they been randomised to JtD intervention. CACE analysis will be performed in the following steps (Peng et al, 2004):

1. Using participants in the JtD group, derive a logistic regression model to predict the probability of being a non-complier (i.e. attending less than 10 of the 16 sessions). Possible predictor covariates are age, sex, type of dementia, baseline QoL.
2. Apply these predictions to the Usual Care group, so that each participant is given a probability of receiving the JtD intervention as planned (if they had been randomised to receive it) which is based on their covariates.
3. For each participant in the Usual Care group calculate a re-weighted outcome defined as the original outcome multiplied by the predicted probability of receiving as planned.
4. Compare the subset of participants in the JtD group that are deemed to have complied with intervention (attending at least 10 sessions) with the re-weighted outcomes amongst

participants in the Usual Care group.

CACE analysis will be conducted by a two stage regression, the first will use logistic regression (excluding clustering adjustment for JtD group) but including site. This is because the clustering is one-sided (only in the JtD arm) so will not add to the predictive model. The second model used in step 4 will be the mixed effects model as used in the primary analysis.

10.3 Mixed effects model checks

Model goodness of fit will be investigated via graphical methods (e.g. histograms of residuals and scatterplots of residuals vs. covariates) for primary and secondary endpoints. Checks will be made on whether a homoscedastic errors assumption is reasonable. Influential observations and outliers will also be investigated and sensitivity analyses at the discretion of the trial statistician will be undertaken and reported.

10.3.1 Clustering Implications and Degrees of Freedoms

First, although this is an individually RCT, part of the JtD intervention (12 of the 16 sessions) is delivered in groups of ideally 8 to 12 participants. Therefore, outcomes of participants who attend group sessions together are more

likely to be similar to each other compared to their counterparts from other group sessions. Furthermore, given that this form of clustering which will happen only in the JtD intervention arm, there are different ways of coding the unclustered Usual Care arm; treating each participant as a single cluster (singletons), all participants as one cluster, or random creation of artificial clusters (Flight, 2016). Another issue is that the appropriate computation of the degrees of freedoms to preserve the type I error is influenced by the number of clusters (Kahan et al., 2016). In summary, statistical models used for the analyses will account for the clustering in one arm of the trial, employ appropriate coding of the unclustered Usual Care arm, and use appropriate approaches to compute degrees of freedoms as supported by current statistical literature.

10.4 Imputation of missing primary outcome data

For sensitivity analysis, imputation will be used in order to obtain complete eight month DEMQOL data. No imputation will be performed for participants that have died before the eight month follow up assessment. Missing data

will be imputed using at least two methods; regression imputation and multiple imputation.

10.4.1 Regression Imputation

Regression imputation will be used to impute missing eight month DEMQOL dimension data. A regression equation that accounts for participant baseline data (age, sex, baseline DEMQOL) will be used to fill in missing values.

10.4.2 Multiple Imputation

One hundred multiple imputation data sets will be created using chained equations. The multiple imputation equation will include baseline data (age, sex, baseline DEMQOL), prognostic factors (type of dementia, medical history) and predictors of missing data (see section 9.5) to make the Missing at Random assumption as plausible as possible. A conservative approach will be adopted and treatment group will be excluded from the imputation model.

11. Data manipulation and definitions

11.1 Definitions

- For the sake of subgroup analysis presence of participating supporter (yes or no) is defined based on the eligibility assessment question “Is there a participating supporter; yes/no”.

- The unblinding of outcome assessors is defined as real when the outcome assessor has correctly identified the allocation group and suspected when the outcome assessor is believed to be unblinded but failed to identify the group allocation.
- Stratification site: Each site is expected to run (at least) two waves of the JtD intervention, potentially at multiple delivery sites. In cases where participants could attend either delivery sites within a site (for example Sheffield) then the stratification site is identical to the site. In cases where sites are recruiting participants from two distinct delivery sites and the participants are only expected to attend an intervention at that delivery site then each delivery site is defined as a stratification site (for example Teas Esk and Wear (the site) is recruiting at both Harrogate and York the delivery sites). The list of stratification sites will be recorded in the SCRAM randomisation system.

11.2 DEMQOL Scoring System

DEMQOL may be collected over the telephone or face-to-face, the scoring and analysis will be consistent regardless of the method of collection but we will note the number of instances of telephone collection. A total score based on the first 28 items (excluding the 29th item) is calculated by summing the items. The score ranges from 28 – 112 and a higher score represents better health related quality of life. For a dementia patient record to be classified as valid, at least 50% of the first 28 items (at least 14 items) must have responses (Smith et al., 2005). Any records with less than 50% completed items will be deemed invalid and excluded from any analysis. The following recommended approach is adopted to deal with missing item data among those with at least 50% item responses (Bsmsacuk, 2016):

- Calculate the participant specific mean based on items with responses
- Impute the missing items with the patient specific mean of completed item responses
- Generate the overall total score by summing all 28 item responses (including imputed data)

11.3 PHQ-9 Scoring System

The PHQ-9 has nine questions; which are assigned values 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. A total score ranging from 0 to 27 is then computed by summing together the assigned scores for the nine questions (Kroenke et al., 2001). The total score can also be categorised with respect to increasing depression: none (0 to 4), mild depression (5 to 9), moderate depression (10 to14), moderately severe depression (15 to 19), and severe depression (20 to 27). A higher score indicates increasing severity of depression. For a PHQ-9 score to be classed as valid at least 7 items should be

complete. The following recommended approach is adopted when two or fewer item are missing (Datadictionary nhs PHQ-9, 2016):

- Calculate the participant specific mean based on items with responses
- Impute the missing items with the patient specific mean of completed item responses
- Generate the overall total score by summing all 9 item responses (including imputed data) •

Round total score to the nearest integer

11.4 GAD-7 Scoring System

For the seven questions; values of 0, 1, 2, and 3 are assigned to the response categories of 'not at all', 'several days', 'more than half the days', and 'nearly every day', respectively. A total score is then computed by summing together the assigned scores for the seven questions (Spitzer et al., 2006). A higher score indicates increasing severity of anxiety, which could also be classified as: no (0 to 4), mild(5 to 9), moderate (10 to 14), and severe anxiety (15 to 21) (Spitzer et al., 2006). The score is valid if at least 5 items are completed. If one or two items are missing, they are substituted by the mean score of the participant non-missing items (Datadictionary nhs GAD-7, 2016).

11.5 SMAS Scoring System

This has a total of 30 questions across 6 subdomains (Talking initiative, Investment behaviour, Variety, Multifunctionality, Self-efficacy, Positive frame of mind). A total score is calculated by summing responses of all 30 items across the six subdomains (Schuurmans et al., 2005). Subdomain scores will not be classed as valid if more than one item is missing. Missing items will be imputed with the mean score of the completed items for that subdomain for the participant (Correspondence with N Steverink, publication pending). The composite total score will be calculated only if all subdomain scores are calculated. The total score ranges from 30 to 175, where higher scores indicate greater self management abilities. Only the total score will be reported.

11.6 GSE Scale Scoring System

All ten questions are assigned values of 1, 2, 3, and 4 to the corresponding categories ‘not true at all’, ‘hardly true’, ‘moderately true’, and ‘exactly true’, respectively. A total score with possible range of 10 to 40 is then calculated by summing all of the 10 questions. For the total score to be valid at least 7 items must be complete. The following recommended approach is adopted in the case of missing data when at least 7 items are complete (Fu-berlinde, 2016): • Calculate the participant specific mean based on items with responses

• Impute the missing items with the patient specific mean of completed item responses

• Generate the overall total score by summing all 10 item responses (including imputed data)

A higher GSE total score indicates more self-efficacy (Schwarzer et al, 1995).

11.7 Diener’s Flourishing Scale Scoring System

A total score based on all 8 items, with varying responses from 1 to 7 is computed to give a possible range of

scores of 8 to 56. The higher the score means that a participant has many psychological resources and strengths. The score will not be calculated if any item responses are missing.

11.8 IADL Scoring System

The IADL consists of 8 items that receive a score of 0 or 1. The total score is calculated by summing the scores for each item, the total score ranges from 0-8. A lower score indicates a higher level of dependence. The IADL will not be calculated if any item responses are missing.

11.9 MMSE Scoring System

The Mini Mental State Examination (MMSE) consists of 11 questions. The total score is calculated by summing the responses. The score ranges from 0-30. A score of 23 or lower is indicative of cognitive impairment. The MMSE will not be calculated if item responses are missing.

11.10 EQ-5D-5L Scoring system

The EQ-5D-5L consists of 5 questions, the score is measured on a scale from –0.22 to 1.00 (good health). The score will not be calculated if any items are missing. The algorithm for scoring the EQ-5D-5L can be found in the Database Specification. The EQ-5D-5L may be collected over the telephone or face-to-face, the scoring and analysis will be consistent regardless of the method of collection but we will note the number of instances of telephone collection.

The EQ-5D VAS your health state today is measured on a scale from 0 (worst imaginable health state) to 100 (best imaginable health state).

11.11 SCQ scoring system

The Sense of Competence Questionnaire score is calculated by summing the answers to the 27 questions. The total score ranges from 27 to 135. A higher score represents a person with a better sense of competence. The SCQ will not be calculated if any questions are missing.

12. Implementation of the SAP

This SAP will be used as a work description for the statistician involved in the trial. All analyses will be

performed by the same statistician (under the supervision of senior trial statistician Professor Stephen Walters) and consequently none of the investigators involved in the trial will perform any of the statistical analyses.

Initially, the data manager will provide blinded data for preliminary checks by the statistician. Due to the nature of the data (clustering in one arm only) it is not feasible for blinded randomisation codes to be released before database lock. Following database freeze, unblinded data will be delivered to the statistician to define analysis sets and test statistical programs. Any queries will be communicated to the study and data manager prior to database lock, and any changes to the database during this time will be documented. The database will be locked after agreement between the statistician, data manager and study manager. It is expected that no data amendments should be required following database lock. However if an amendment is required, the process is documented in CTRU SOP DM012.

13.Questionnaire summary table

Name	No. of items	Score range	Description	Interpretation of score
DEMQOL	29	28-112	Measures health related Qol in people with dementia	self-efficacy higher DEMQOL score means higher health related quality of life
PHQ-9	9	0-27	Measures severity of depression	higher score indicates increasing severity of depression
GAD-7	7	0-21	is a severity measure of generalised anxiety disorder	higher score indicates increasing severity of anxiety
IALD	8	0-8	measures independent living skills	higher score indicates lower level of dependence
GSE	10	10-40	measures general perceived self efficacy	higher score indicates more self efficacy
Diener's Flourishing scale	8	56	measures psychological flourishing	higher score means participant has many psychological resources and strengths
SMAS	30	30-175	measures self management abilities	higher SMAS indicates greater self management abilities
SCQ	27	27-135	measures sense of competence in caregivers	higher score represents a person with a better sense of competence
EQ-5D-5L value index	5	(-0.224, 1)	Measure of health status. 5 domains include mobility, self care, usual activities, pain/discomfort, and anxiety/depression.	A score of zero means death, 1 is full health, negative score is a state worse than death
EQ-5D-5L VAS	1	(0, 100)	Measure of health status.	A score of zero means worst health and 100 means best health.

14. Example Tables and Figures

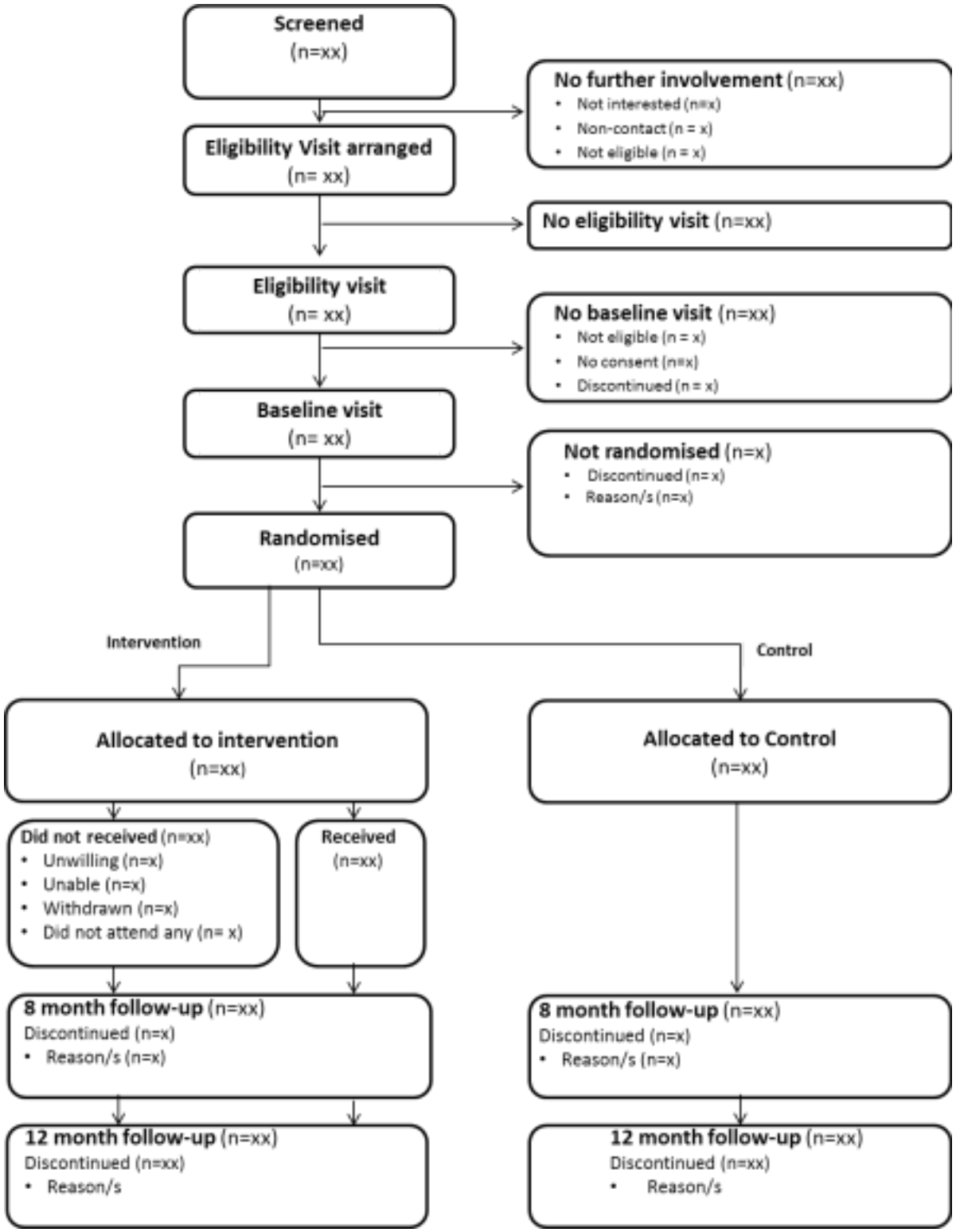

Figure 1 CONSORT flow diagram of participants through the trial

Table 4. Socio-demographic and characteristics of participants.

Variable Scoring Usual Care JtD All (n=xx) (n=xx) (N=xx)

Delivery site Sheffield xx(xx%) xx(xx%) xx(xx%) Bradford xx(xx%) xx(xx%) xx(xx%)
Leicester xx(xx%) xx(xx%) xx(xx%)
Hull xx(xx%) xx(xx%) xx(xx%)
York xx(xx%) xx(xx%) xx(xx%)
Halifax xx(xx%) xx(xx%) xx(xx%)
Nottingham xx(xx%) xx(xx%) xx(xx%)
Leeds xx(xx%) xx(xx%) xx(xx%)
Oldham xx(xx%) xx(xx%) xx(xx%)

Sex Male xx(xx) xx(xx) xx(xx) Female xx(xx) xx(xx) xx(xx)

Age (years) Mean (SD) xx(xx) xx(xx) xx(xx) Median (IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

Ethnicity [a] White [b] xx(xx%) xx(xx%) xx(xx%) Mixed/multiple ethnic groups [c] xx(xx%) xx(xx%) xx(xx%)
Asian/Asian British [d] xx(xx%) xx(xx%) xx(xx%)
Black/African/Caribbean/Black British [e] xx(xx%) xx(xx%) xx(xx%)

Other ethnic group [f] xx(xx%) xx(xx%) xx(xx%)
Prefer not to say xx(xx%) xx(xx%) xx(xx%)

Living with others Yes xx(xx%) xx(xx%) xx(xx%)

Live with (only yes above) Spouse/partner xx(xx%) xx(xx%) xx(xx%) Child/children xx(xx%) xx(xx%) xx(xx%)
Parent(s) xx(xx%) xx(xx%) xx(xx%)
Other xx(xx%) xx(xx%) xx(xx%)

[a] Main ethnic groups could be collapsed depending on the observed distribution. [b] White: English/Welsh/Scottish/Northern Irish/British, Irish, Gypsy or Irish Traveller, and Any other White background; [c] Mixed/multiple ethnic groups: White and Black Caribbean, White and Black African, White and Asian, and Any other mixed/multiple ethnic groups background; [d] Asian/Asian British: Indian, Pakistani, Bangladeshi, Chinese, and Any other Asian background; [e] Black/African/Caribbean/Black British: African, Caribbean, and Any other Black/African/Caribbean/Black British background; [f] Other ethnic group: Arab, and Any other ethnic group.

Table 5: Baseline medical history of participants.

Variable Scoring Usual Care JtD All (n=xx) (n=xx) (N=xx)
Specific health problems Stroke xx(xx%) xx(xx%) xx(xx%) Diabetes xx(xx%) xx(xx%) xx(xx%)
Heart or chest problems xx(xx%) xx(xx%) xx(xx%)
Arthritis xx(xx%) xx(xx%) xx(xx%)
Sensory impairment xx(xx%) xx(xx%) xx(xx%)
Falls/dizziness/blackouts xx(xx%) xx(xx%) xx(xx%)
Any other xx(xx%) xx(xx%) xx(xx%)

Type of dementia diagnosed Alzheimer's xx(xx%) xx(xx%) xx(xx%) Vascular dementia xx(xx%) xx(xx%) xx(xx%)
Mixed Alzheimer's/ Vascular dementia xx(xx%) xx(xx%) xx(xx%)

Other xx(xx%) xx(xx%) xx(xx%) Not known xx(xx%) xx(xx%) xx(xx%)

Table 6: Baseline characteristics and quality of life of participants for the ITT set.

Variable Scoring Usual Care JtD All (n=xx) (n=xx) (N=xx)
MMSE (total score) [a](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Media(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)

Min to Max xx to xx xx to xx xx to xx
MMSE cognitive impairment Mild (18 to 23) ** xx(xx%) xx(xx%) xx(xx%) Normal
(n=xx) (n=xx) (N=xx) (24 to 30) xx(xx%) xx(xx%) xx(xx%)

DEMQOL (total score) [b](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to max xx to xx xx to xx xx to xx
DEMQOL Quality of life overall (Q29) xx(xx%) xx(xx%) Fair xx(xx%) xx(xx%) xx(xx%)
(n=xx) (n=xx) (N=xx) Poor xx(xx%) xx(xx%) xx(xx%)
Very good xx(xx%) xx(xx%) xx(xx%) Good xx(xx%)

PHQ-9 (total score) [c](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

PHQ-9 Depression severity None (0 to 4) xx(xx%) xx(xx%) xx(xx%) Mild (5 to 9) xx(xx%) xx(xx%)
xx(xx%)
Moderate (10 to14) xx(xx%) xx(xx%) xx(xx%)
Moderately severe (15 to 19) xx(xx%) xx(xx%) xx(xx%)

Severe (20 to 27) xx(xx%) xx(xx%) xx(xx%)

GAD-7 (total score) [d](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

GAD-7 anxiety severity Mild (0 to 5) xx(xx%) xx(xx%) xx(xx%) Moderate (6 to10) xx(xx%) xx(xx%)
xx(xx%)
Moderately severe (11 to 15) xx(xx%) xx(xx%) xx(xx%)

Severe (16 to 21) xx(xx%) xx(xx%) xx(xx%)

IADL (total score) [e](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

GSE (total score) [f](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

Diener's Flourishing Scale (total score) [g] to xx) xx(xx to xx) xx(xx to xx) Min to Max xx to xx
(n=xx) (n=xx) (N=xx) xx to xx xx to xx
Mean(SD) xx(xx) xx(xx) xx(xx) Median(IQR) xx(xx

SMA [h](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

EQ-5D-5L (value index)[i](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

EQ-5D VAS[i](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

[a]indicator of cognitive impairment [b] higher mean DEMQOL total scores means higher HRQoL; [c] a higher PHQ-9 total score indicates increasing severity of depression [d] a higher GDA-7 total score indicates increasing severity of anxiety [e] a lower IADL total score indicates a higher level of dependence; [f] higher GSE total score indicates more self-efficacy; [g] higher flourishing score represents a participant with many psychological

resources and strengths; [h] a higher SMA composite total score relates to more positive well-being; [i]higher EQ-5D represents better health.

Table 7: Baseline characteristics of participating supporters.

Variable Scoring <u>Usual Care JtD All</u> (n=xx) (n=xx) (N=xx)

Relationship with participants (n=xx) (n=xx) (N=xx) Spouse xx(xx%) xx(xx%) xx(xx%)
Child xx(xx%) xx(xx%) xx(xx%)
Sibling xx(xx%) xx(xx%) xx(xx%)
Other family member xx(xx%) xx(xx%) xx(xx%)

Friend xx(xx%) xx(xx%) xx(xx%)
Neighbour xx(xx%) xx(xx%) xx(xx%)
Paid Carer xx(xx%) xx(xx%) xx(xx%)
Other xx(xx%) xx(xx%) xx(xx%)

Live with participants Yes xx(xx%) xx(xx%) xx(xx%) participants' first choice for support Yes xx(xx%) xx(xx%) xx(xx%)

Sex Male xx(xx) xx(xx) xx(xx) Female xx(xx) xx(xx) xx(xx)

Age (years) Mean (SD) xx(xx) xx(xx) xx(xx) Median (IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

SCQ (total score) [a](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to max xx to xx xx to xx xx to xx

PHQ-9 (total score) [b](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

EQ-5D-5L (value index) [c](n=xx) (n=xx) (N=xx) Mean(SD) xx(xx) xx(xx) xx(xx)
Median(IQR) xx(xx to xx) xx(xx to xx) xx(xx to xx)
Min to Max xx to xx xx to xx xx to xx

a),a higher SCQ score represents a better sense of competence; b) a higher PHQ-9 total score indicates increasing severity of depression: c) a higher EQ-5D-5Lscore represents better health.

Table 8. Exploratory effect of JtD intervention by subgroups: DEMQOL at 8 months.

Subgroup Classification <u>Usual Care JtD</u> Adjusted mean P

<u>(95% CI)</u> [a]

Type of dementia Alzheimer's xx xx(xx) xx xx(xx) xx (xx to xx) Vascular dementia xx xx(xx) xx xx(xx) xx (xx to xx)

value [b] n Mean(SD) n Mean(SD)
difference

Mixed Alzheimer's/ Vascular dementia xx xx(xx) xx xx(xx) xx (xx to xx)
<u>Other xx xx(xx) xx xx(xx) xx (xx to xx) x.xxx</u>

Presence of supporting carer No xx xx(xx) xx xx(xx) xx (xx to xx) Yes xx xx(xx) xx xx(xx) xx (xx to xx) x.xxx

Usual care as the reference group; [a] adjusted for baseline DEMQOL (total score) and stratification site; [b] Overall interaction test. Higher mean DEMQOL total scores means higher HRQoL; SD=standard deviation; CI=Confidence Interval.

Table 9: Continuous baseline characteristics by group and missing data status

Variable Summary	Statistic	Completers			Non-completers		
		Usual Care (n=XX)	JtD (n=XX)	All (n=XX)	Usual Care (n=xx)	JtD (n=xx)	All (n=xx)
Age (yrs)	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
MMSE	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
DEMQoL	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)

PHQ-9 Mean(SD) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) Median(IQR) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x)

GAD-7 Mean(SD) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) Median(IQR) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x)

IADL Mean(SD) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) Median(IQR) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x)

GSE Mean(SD) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) Median(IQR) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x)

EQ-5D-5L (value index)

Diener's
Flourishing Scale
Mean(SD) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) Median(IQR) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x)

Mean(SD) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) xx.x(xx.x) Median (IQR) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x) xx.x(xx.x to xx.x)

Table 10: Categorical baseline characteristics by intervention group and missing data status.

Variable Scoring <u>Completers Non-completers</u> Usual Care JtD All Usual Care JtD All (n=xx) (n=xx) (n=xx) (n=xx) (n=xx) (n=xx)
Sex Male xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) Female xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%)

Type of dementia diagnosed
Alzheimer's xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) Vascular dementia xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) Mixed Alzheimer's/ Vascular dementia xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) Other xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%)

Live with others Yes xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) xx(xx.x%) … … … … … .. … …

Table 11. Primary effectiveness analysis: DEMQOL total score at 8 months.

	Primary outcome Usual Care JtD		Adjusted mean P-value [b]	
	n Mean(SD)	n Mean(SD)	Adjusted mean difference (95% CI) [a]	P value [a]
DEMQOL total score	xx xx(xx)	xx xx(xx)	xx (xx to xx)	x.xxx xx(xx to xx) x.xxx

Usual care as the reference group; [a] adjusted for baseline DEMQOL (total score) and stratification site; [b] adjusted for baseline DEMQOL total score, stratification site, and additional covariates; age, sex, PHQ-9 (total score), and GAD-7 (total score). Higher mean DEMQOL total scores means higher HRQoL.

Table 12. Sensitivity analyses on the primary outcome: DEMQOL total score at 8 months.

Primary outcome: DEMQOL total score	Usual Care JtD n Mean(SD)	n Mean(SD)	Adjusted mean difference (95% CI) [a]	P value [a]	Adjusted mean difference [b] (95% CI)	P value [b]
MI	xx xx(xx)	xx xx(xx)	xx (xx to xx)	x.xxx	xx(xx to xx)	x.xxx
Regression imputation	xx xx(xx)	xx xx(xx)	xx (xx to xx)	x.xxx	xx(xx to xx)	x.xxx
CACE	xx xx(xx)	xx xx(xx)	xx (xx to xx)	x.xxx	xx(xx to xx)	x.xxx
MM	xx xx(xx)	xx xx(xx)	xx (xx to xx)	x.xxx	xx(xx to xx)	x.xxx

Usual care as the reference group; [a] adjusted for baseline DEMQOL (total score) and stratification site; [b] adjusted for baseline DEMQOL total score, stratification site, and additional covariates; age, sex, PHQ-9 (total score), and GAD-7 (total score). Higher mean DEMQOL total scores means higher HRQoL; SD=standard deviation, CI=Confidence Interval, CACE=Complier Average Causal Effect, MVI=Mean Value Imputation, MM=Mistimed Measurements.

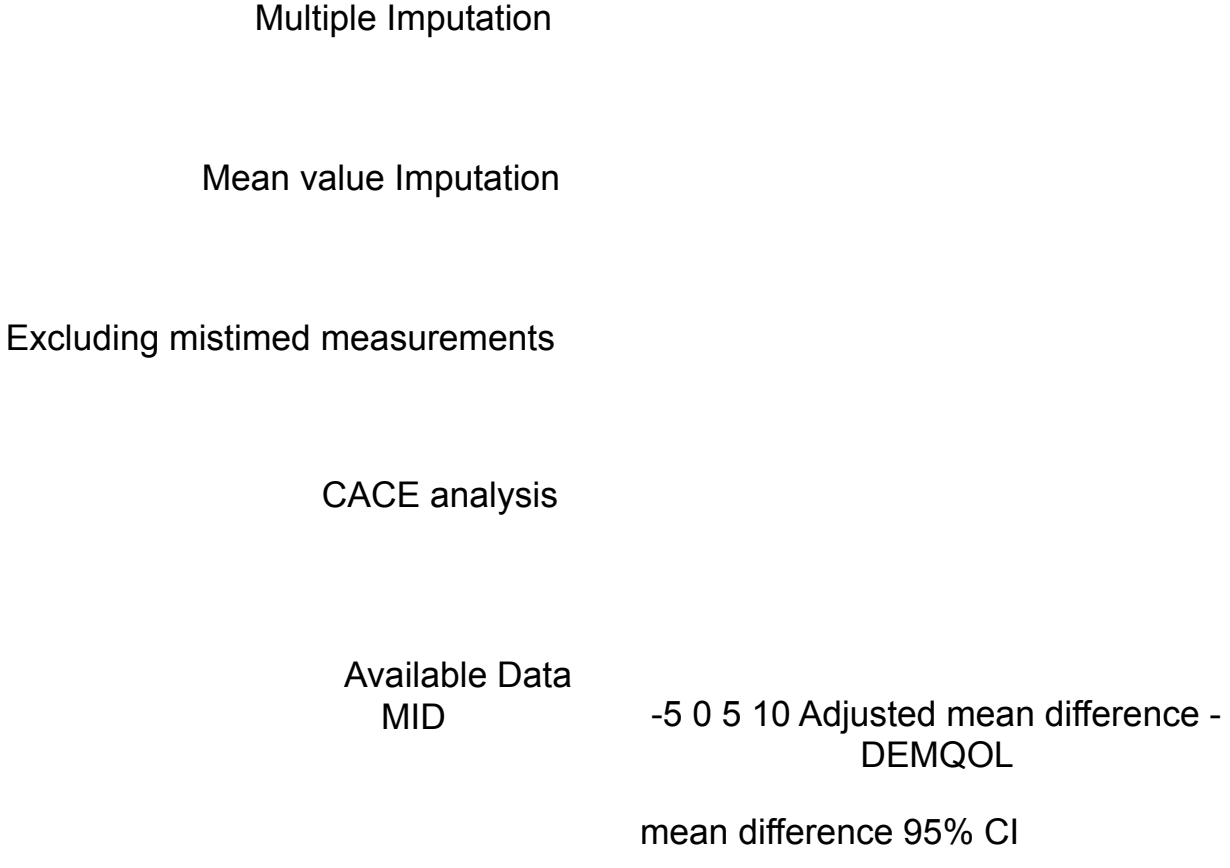

Figure 2 Forest plot of sensitivity analysis on primary outcome: DEMQOL at 8 months

Table 13. Exploratory effect of JtD intervention by subgroups: DEMQOL at 8 months.

Subgroup Classification	**Usual Care**		**JtD**		**Adjusted mean difference (95% CI)** [a]	**P value** [b]
	n	**Mean(SD)**	**n**	**Mean(SD)**		
Type of dementia						
Alzheimer's	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	
Vascular dementia	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	
Mixed Alzheimer's/ Vascular dementia	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	
Other	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
Presence of supporting carer						
No	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	
Yes	xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx

Usual care as the reference group; [a] adjusted for baseline DEMQOL (total score) and stratification site; [b] Overall interaction test. Higher mean DEMQOL total scores means higher HRQoL; SD=standard deviation; CI=Confidence Interval.

Table 14. Effectiveness analysis: secondary outcomes at 8 and 12 months

Secondary outcome and timing	**Usual Care**		**JtD**		**Adjusted mean difference (95% CI)** [a]	**P-value** [a]
	n	**Mean(SD)**	**n**	**Mean(SD)**		
At 8 Months						
PHQ-9 (total score)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
GAD-7 (total score)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
IADL (total score)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
GSE (total score)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
Diener's Flourishing Scale (total score)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
SMA Composite total score	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
EQ-5D-5L (value index)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
EQ-5D-5L (VAS)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
At 12 months						
DEMQOL (total score)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
EQ-5D-5L (value index)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx
EQ-5D-5L (VAS)	Xx	xx(xx)	xx	xx(xx)	xx (xx to xx)	x.xxx

Usual care as the reference group; [a] adjusted for baseline outcome measure under consideration and stratification site using a mixed effects linear regression model; SD=standard deviation; CI=Confidence Interval; Higher mean DEMQOL total scores means higher HRQoL; Higher mean flourishing score represents a participant with many psychological resources and strengths; a higher PHQ-9 total score indicates increasing severity of depression, a higher GDA-7 total score indicates increasing severity of anxiety · a lower IADL total score indicates a higher level of dependence; higher GSE total score indicates more self-efficacy; higher flourishing score represents a participant with many psychological resources and strengths; a higher SMA composite total score relates to more positive well-being; [i]higher EQ-5D-5L represents better health.

Table 15. Serious Adverse Events by treatment group

Serious Adverse Events	Intervention (n=XXX)	Control (n= XXX)	All (n= XXX)

Number (%) of participants who experienced ≥1 SAE XXX (xx%) XXX (xx %) XXX (xx %) Number

of all SAEs (including repeated events) XXX XXX XXX

Seriousness
Death xx (xx%) xx (xx%) xx (xx%) Life threatening xx (xx%) xx (xx%) xx (xx%) Inpatient hospitalisation xx (xx%) xx (xx%) xx (xx%) Prolongs hospitalisation xx (xx%) xx (xx%) xx (xx%) Persistent or significant disability/incapacity xx (xx%) xx (xx%) xx (xx%) *Total* xx (xx%) xx (xx%) xx (xx%)

Intensity
Mild xx (xx%) xx (xx%) xx (xx%) Moderate xx (xx%) xx (xx%) xx (xx%) Severe xx (xx%) xx (xx%) xx (xx%) *Total* xx (xx%) xx (xx%) xx (xx%)

Relationship to study intervention

Unlikely xx (xx%)
Unrelated xx (xx%)
Total xx (xx%)

Table 16: Harms by treatment group: GAD-7 and PHQ-9

	Intervention (n=XXX)	Control (n= XXX)	All (n= XXX)
Number with symptoms of anxiety (GAD-7≥8)			
Baseline	XXX (xx%)	XXX (xx %)	XXX (xx %)
8 months	XXX (xx%)	XXX (xx %)	XXX (xx %)
12 months	XXX (xx%)	XXX (xx %)	XXX (xx %)
Number with symptoms of depression (PHQ-9≥10)			
Baseline	XXX (xx%)	XXX (xx %)	XXX (xx %)
8 months	XXX (xx%)	XXX (xx %)	XXX (xx %)
12 months	XXX (xx%)	XXX (xx %)	XXX (xx %)

15.References

Abraha, I., Cherubini, A., Cozzolino, F., De Florio, R., Luchetta, M.L., Rimland, J.M., Folletti, I., et al. (2015), “Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study.”, *BMJ (Clinical Research Ed.)*, Vol. 350, p. h2445.

Altman, D. G. (1990). Practical statistics for medical research. CRC press.

Assmann, S.F., Pocock, S.J., Enos, L.E. and Kasten, L.E. (2000), “Subgroup analysis and other (mis)uses of baseline data in clinical trials.”, *Lancet (London, England)*, Vol. 355 No. 9209, pp. 1064–9.

Baldwin, S.A., Bauer, D.J., Stice, E. and Rohde, P. (2011), “Evaluating models for partially clustered designs.”, *Psychological*

Methods, Vol. 16 No. 2, pp. 149–65.

Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. (2008) Methods and Processes of the CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments. Ann Intern Med. W60-W67. PMID: 18283201

Bsmsacuk. (2016). Bsmsacuk. Retrieved 15 November, 2016, from https://www.bsms.ac.uk/research/cds/research/demqol.aspx

Brazier J, Ratcliffe J, Salomon J & Tsuchiya A (2007) Measuring and Valuing Health Benefits for Economic Evaluation. 1st Ed. Oxford: Oxford University Press.

Datadictionary nhs PHQ-9. (2016). Datadictionarynhsuk. Retrieved 15 November, 2016, from http://www.datadictionary.nhs.uk/data_dictionary/nhs_business_definitions/p/patient_health_questionnaire 9_de.asp?shownav=1

Datadictionary nhs GAD-7. (2016). Datadictionarynhsuk. Retrieved 15 November, 2016, from http://www.datadictionary.nhs.uk/data_dictionary/nhs_business_definitions/g/generalised_anxiety_disorder_questionnaire_d e .asp?shownav=1

de Boer, M.R., Waterlander, W.E., Kuijper, L.D.J., Steenhuis, I.H.M. and Twisk, J.W.R. (2015), "Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate.", *The International Journal of Behavioral Nutrition and Physical Activity*, Vol. 12 No. 1, p. 4.

Cuzick, J. (2005), "Forest plots and the interpretation of subgroups.", *Lancet (London, England)*, Vol. 365 No. 9467, p. 1308.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. Social Indicators Research, 97(2), 143-156.

Flight, L., Allison, A., Dimairo, M., Lee, E., Mandefield, L., & Walters, S. J. (2016). Recommendations for the analysis of individually randomised controlled trials with clustering in one arm–a case of continuous outcomes. BMC Medical Research Methodology, 16(1), 165.

Fu-berlinde. (2016). Fu-berlinde. Retrieved 15 November, 2016, from http://userpage.fu-berlin.de/~health/faq_gse.pdf

Gold, D. A. (2012). An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. Journal of clinical and experimental neuropsychology, 34(1), 11-34.

ICH. (1998), *ICH E9: Guidance on Statistical Principles for Clinical Trials*, available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf.

Hancock, P., & Larner, A.J., (2009) "Clinical utility of Patient Health Questionnaire-9 (PHQ-9) in memory clinics" .Int J Psychiatry

ClinPract; Vol. 13 No. 3, p.188-91.

Kahan, B.C., Forbes, G., Ali, Y., Jairath, V., Bremner, S., Harhay, M.O., Hooper, R., et al. (2016), “Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study”, *Trials*, BioMed Central, Vol. 17 No. 1, p. 438.

Knol, M., Groenwold, R. and Grobbee, D. (2012), “P-values in baseline tables of randomised controlled trials are inappropriate but still common in high impact journals”, *European Journal of Preventive Cardiology*, Vol. 19 No. 2, pp. 231–232.

Kroenke, K., Spitzer, R.L. and Williams, J.B. (2001), “The PHQ-9: validity of a brief depression severity measure.”, *Journal of General Internal Medicine*, Springer, Vol. 16 No. 9, pp. 606–13.

Mountain, G., Windle, G., Hind, D., Walters, S.,, Keertharuth, A., Chatters, R., Sprange, K., Craig, C., Cook, S., Lee, E, …, Roberts, J.; (2017) A preventative lifestyle intervention for older adults (lifestyle matters): a randomised controlled trial. *Age Ageing 1-8.* doi: 10.1093/ageing/afx021

Mulhern, B., Rowen, D., Brazier, J., Smith, S., Romeo, R., Tait, R., ... & Lamping, D. (2013). Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. Health technology assessment (Winchester, England), 17(5), 1-+.

Pocock, S.J., Assmann, S.E., Enos, L.E. and Kasten, L.E. (2002), “Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems.”, *Statistics in Medicine*, Vol. 21 No. 19, pp. 2917–30.

Peng, Y., Little, R. J., & Raghunathan, T. E. (2004). An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics*, *60*(3), 598-607.

Roberts, C., Batistatou, E. and Roberts, S.A. (2015), “Design and analysis of trials with a partially nested design and a binary outcome measure”, *Statistics in Medicine*, No. April 2014, p. n/a-n/a.

Satterthwaite, F.E. (1946), “An approximate distribution of estimates of variance components.”, *Biometrics*, Vol. 2 No. 6, pp. 110–4.

Schuurmans, H., Steverink, N., Frieswijk, N., Buunk, B.P., Slaets, J.P.J. and Lindenberg, S. (2005), “How to measure self management abilities in older people by self-report. The development of the SMAS-30”, *Quality of Life Research*, Kluwer Academic Publishers, Vol. 14 No. 10, pp. 2215–2228.

Schwarzer. R., & Jerusalem. A., (1995) Generalized Self-Efficacy scale. In J.Weinman, S. Wright &M.Johnston. Measures in health psychology: A user’s portfolio. Causal and control beliefs; 35-57. Windsor, UK: NFER-NELSON.

Senn, S. (1994), “Testing for baseline balance in clinical trials.”, *Statistics in Medicine*, Vol. 13 No. 17, pp. 1715–26.

Smith, S.C., Lamping, D.L., Banerjee, S., Harwood, R., Foley, B., Smith, P., Cook, J.C., et al. (2005), “Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of

current methodology.", *Health Technology Assessment (Winchester, England)*, Vol. 9 No. 10, pp. 1–93, iii–iv.

Spitzer, R.L., Kroenke, K., Williams, J.B.W. and Löwe, B. (2006), "A brief measure for assessing generalized anxiety disorder: the GAD-7.", *Archives of Internal Medicine*, Vol. 166 No. 10, pp. 1092–7.

Walters, S.J. (2010), "Therapist effects in randomised controlled trials: what to do about them.", *Journal of Clinical Nursing*, Vol. 19 No. 7–8, pp. 1102–12.

Wang, R., Lagakos, S.W., Ware, J.H., Hunter, D.J. and Drazen, J.M. (2007), "Statistics in medicine--reporting of subgroup analyses in clinical trials.", *The New England Journal of Medicine*, Massachusetts Medical Society, Vol. 357 No. 21, pp. 2189–94.

Wild, B., Eckl, A., Herzog, W., Niehoff, D., Lechner, S., Maatouk, I., ... & Löwe, B. (2014). Assessing generalized anxiety disorder in elderly people using the GAD-7 and GAD-2 scales: results of a validation study. *The American Journal of Geriatric Psychiatry*, *22*(10), 1029-1038.

Vernooij-Dassen, M., Kurz, X., Scuvee-Moreau, J., & Dresse, A. (2003). [The measurement of sense of competence in caregivers of patients with dementia]. *Revue d'epidemiologie et de sante publique*, *51*(2), 227-235.