ISSN 2399-3502

University of Nottingham
UK | CHINA | MALAYSIA

Nottingham University Hospitals NHS
NHS Trust

www.nottingham.ac.uk/praised/index.aspx

PrAISED
Promoting Activity, Independence and Stability in Early Dementia

The Promoting Activity, Independence and Stability in Early Dementia (PrAISED) research programme is a NIHR funded project that has been designed to help people with mild cognitive impairment or early stage dementia to remain healthier and more independent for longer. We have designed an activity and exercise programme consisting of a combination of exercises, activities of daily living and memory strategies to help improve and maintain individual physical and mental health.

PrAISED Discussion Paper Series
ISSN 2399-3502

Issue 6, November 2022

Statistical Analysis Plan for the Promoting Activity, Independence and Stability In Early Dementia and Mild Cognitive Impairment (PrAISED) Full-Scale RCT

Brand A[1], Hoare Z[1], Harwood RH[2,3]

Author Affiliations:
1. North Wales Organisation for Randomised Trials in Health & Social Care (NWORTH), School of Medical and Health Sciences, Bangor University, UK
2. School of Health Sciences, University of Nottingham, UK
3. Nottingham University Hospitals, UK

Address for Correspondence:
Dr Andrew Brand, NWORTH, School of Medical and Health Sciences, Bangor University, Y Wern, Normal Site, Holyhead Road, Bangor, Gwynedd, LL57 2PZ. Email: a.brand@bangor.ac.uk

TABLE OF CONTENTS

1. ACRONYMS AND DEFINITION OF TERMS

Acronym	Meaning
ADL	Activities of Daily Living
AE	Adverse Event
ANCOVA	Analysis of Covariance
CANTAB	Cambridge Automated Neuropsychological Testing Battery
CHEME	Centre for Health Economics and Medicines Evaluation
CSI	Carer Strain Index
CSRI	Client service receipt inventory
DAD	Disability Assessment for Dementia
DEMQOL	Dementia quality of life
DMEC	Data Monitoring & Ethics Committee
DMP	Data Management plan
EQ-5D	EuroQol five dimensions questionnaire
FES	Falls Efficacy Scale
HADS	Hospital anxiety and depression scale
ITT	Intention to Treat
LAPAQ	LASA (Longitudinal Aging Study Amsterdam) Physical Activity Questionnaire (LAPAQ)
NWORTH	North Wales Organisation for Randomised Trials in Health
PrAISED	Promoting Activity, Independence and Stability in Early Dementia
PSC	Programme Steering Committee
RCT	Randomised Controlled Trial
SAE	Serious Adverse Event
SOP	Standard Operating Procedure
SPSS	Statistics Package for the Social Sciences
TUG	Timed up and go test

3. INTRODUCTION

3.1 BACKGROUND AND RATIONALE

People with dementia are at high risk of decline in physical and mental health, functional ability and independence. To address this problem a therapy programme has been developed (PrAISED), which promotes activity and reduces risk of falls, helping people to live well with dementia for longer. In order to assess the clinical and the cost effectiveness of the PrAISED therapy intervention a multi-centre randomised controlled trial has been conducted. The trial protocol has been published, see Bajwa R et al (2019).

3.2 STUDY OBJECTIVES

PURPOSE

To determine the clinical and cost-effectiveness of a newly developed therapy programme to promote activity and independence, and prevent falls, suitable for people with early dementia and mild cognitive impairment (PrAISED).

OBJECTIVES

Primary Objective

- To determine if the PrAISED intervention reduces disability in Activities of Daily Living measured by the DAD at the 12 months follow-up.

Secondary Objectives

To, additionally, determine if the PrAISED intervention at the 12 months follow-up:

- Decreases rate of falling and increases the time to first fall (months 3-15).
- Improves quality of life.
- Increases level of habitual activity.
- Improves cognition.
- Reduces the rate of fractures and injurious falls (months 3-15).
- Improves balance and functional mobility.
- Improves mood, anxiety and fear of falling.
- Reduces frailty.

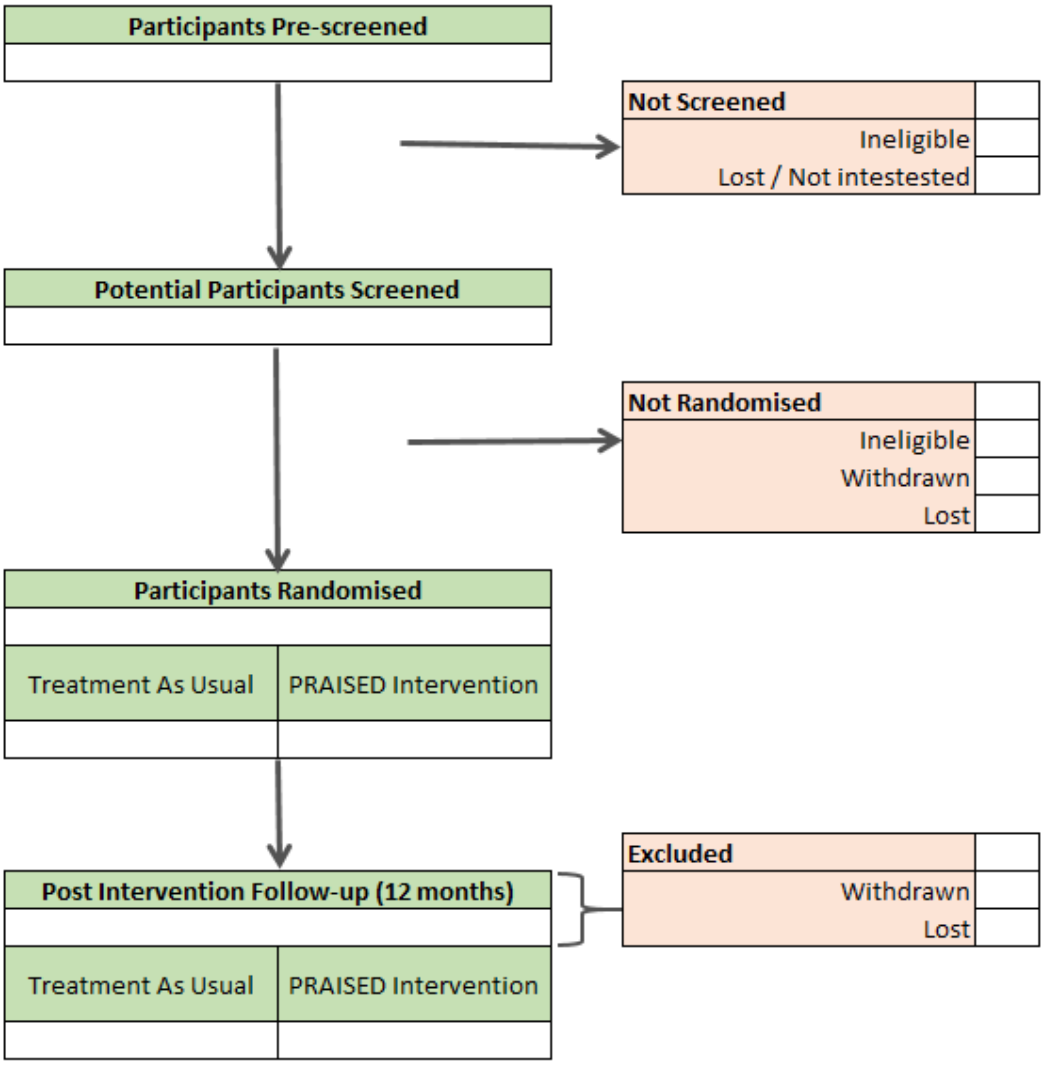
- Reduces the rate of hospital and care home admissions, and days spent in hospital.
- Reduces carer strain.
- Reduces apathy.
- Is cost-effective, within the trial period, over the anticipated remaining lifespan, and using a social return on investment model.

3.3 TRIAL DESIGN

The trial is a multi-centred randomised controlled trial (RCT). The RCT involves an individually randomised, parallel group design with two groups PrAISED therapy programme (the intervention condition) or brief falls prevention assessment and prevention advice (the treatment as usual condition). A process evaluation and health economic analysis will also be conducted. The process evaluation is described by Di Lorito et al (2019).

3.4 CONSORT DIAGRAM

Figure 1. CONSORT flowchart template



4. STATISTICAL PRINCIPLES

4.1 SAMPLE SIZE JUSTIFICATION

A sample size of 184 participants per group, taking into account 23% attrition (informed by the data from previous studies conducted by Blankevoort 2010 and Pitkala 2013), has 80% power to detect changes in the disability outcome, DAD, with effect size 0.5 (11 points on a baseline of 70, standard deviation 22, data from [Blankevoort 2010, Rao 2014]) and an adjustment to allow for underestimation of

the population standard deviation (i.e., the sample size for each group was multiplied by 2.25). In the feasibility study we observed an attrition rate of 18.4% and a standard deviation of 21 on the DAD outcome measure. However, the DMEC committee recommend that we err on the side of caution and continued with the originally planned sample size of 368.

The trial was disrupted by the COVID-19 pandemic, but recruitment was completed following extension, mitigations and changes to recruitment, intervention delivery and follow up
procedures. Recruitment closed on 30th June 2021 with 365 participants recruited.

4.2 RANDOMISATION

Randomisation was performed by NWORTH using a secure, web-based system that could be accessed 24-hours a day. The system uses a dynamic allocation algorithm devised by Russel, Hoare, Whitaker, Whitaker and Russell (2011). This method protects against subversion while ensuring that the trial maintains good balance to the allocation ratio of 1:1, both within each stratification variable and across the trial. Participants were randomised to brief falls assessment and prevention advice or the PrAISED Therapy intervention group. Participants will be stratified by:

1. Site: Nottingham, Derby, Lincoln, Oxford and Bath.
2. Having a co-resident carer: Yes or No.
3. History of previous falls (i.e., a participant has had one or more falls in the last year): Yes or No.

4.3 LEVELS OF CONFIDENCE AND P-VALUES

All statistical tests are two-sided and performed using a 5% significance level. Similarly, all presented confidence intervals are 95% and two-sided. P values, effect size estimates and 95% confidence intervals will be reported for all analyses. Interpretation of results from analyses will primarily focused on the effect size estimate and its associated 95% confidence intervals. In line with the recommendations of Althouse (2016) and Li et al (2017), analyses of secondary outcomes and exploratory analyses, p-values and confidence intervals will *not* be adjusted to take into account multiple comparisons.

The rationale for comparing multiple outcomes in this way is:

1. The intervention is complex and is likely to affect a range of physical health, functional, psychological, and social outcomes. It will not be possible to anticipate the relative impacts on different outcomes in advance.
2. We are interested in both impact on health status and the mechanism of impact
3. Intervention effects might credibly vary (interact) with baseline variables.
4. We will rely on consistency of effect across different outcome variables, for example displayed by a forest plot, as much as p-values in determining the play of chance.

4.4 ADHERENCE

A descriptive approach will be taken with regard to treatment adherence that will focus on the number and duration of therapy sessions completed (defined as session commenced and any activity undertaken) for each participant, and the frequency and duration that prescribed exercises were undertaken outside of therapy sessions (ascertained from participant calendars). This is being studied in the Process Evaluation.

4.5 WITHDRAWALS AND MISSING DATA

Missing data rules are considered an integral part of a validated measurement tool. Therefore, for missing items within an outcome measure, the published rules for completing missing data for the relevant measure will be applied. If there are no missing data rules for an outcome measure the following methods will be employed in the following order.

In line with the recommendation of Bono, Ried, Kimberlin and Vogel (2007), if the number of missing items on an outcome measure is 20% or less, then the missing value for the item will be substituted by the either the individual's mean score for the remaining items on the scale or the mean score for the item. The substitution employed will be dependent upon the nature of the outcome measure. If there are more than 20% missing items in the scale the outcome measure will not be calculated for the participant at that time point. Note though that this approach may not be appropriate for all outcome measures.

For each outcome measure, the proportion of participants without an outcome measure score will be calculated. This will be done for the both the baseline assessment and 12 months follow-up.

In accordance with the recommendations of Jakobsen, Gludd, Wetterslev and Winkel (2017), if the proportion of missing data is less than 5%, no imputation of missing data will be performed, and analyses will be based on complete cases. If the proportion of missing data is equal to or greater the 5%, a missing completely at random test devised by Little (1998) will be performed to assess whether the data is missing completely at random (MCAR). If this test indicates that data is missing completely at random then analyses will be based on complete cases, otherwise independent t-tests and Chi-square tests will be conducted to investigate whether the data is missing at random (MAR). If these tests suggest that the missing data is MAR the predictive mean matching multiple imputation method will be employed. Otherwise, additional modelling guided by clinical knowledge would be required to simulate the missing data mechanism and impute the missing data.

For multiple imputations, the number of imputations completed will be dependent upon the percentage of missing data (White, Royston, & Wood, 2011). The missing outcome measures at baseline will be imputed using all the stratification factors and the participant's (or carer's, for the carer measures) baseline demographics. The missing outcome measures for the 12 months follow-up will be imputed using all the stratification variables, participant's (or carer's, for the carer measures) baseline demographics and the scores at baseline. In line with the recommendations of Jakobsen et al (2017), if more than 40% of data is missing for an outcome measure the data will not be imputed and an analysis will be conducted on complete cases, though the results of the analysis will be interpreted tentatively.

Ad-hoc interim follow-up data (see section 7.1 for details) may also be used as a substitute for missing 12 months follow-up, if the Ad-hoc interim follow-up occurred at least 6 months after the baseline assessment.

Descriptive statistical analyses of the withdrawals between the treatment as usual and PrAISED intervention groups will be conducted. These analyses will focus on the baseline characteristics of the participant and the nature of the withdrawal (i.e., meaningful [e.g., new care home placement, death or deterioration in health or cognition] or non-meaningful [e.g., participant choice or no reason]).

4.6 OUTLIERS

Outliers identified from the statistical analyses will be examined by rechecking the data. This will involve running a test for outliers proposed by Grubbs (1969) and visually inspecting a boxplot. No outliers will be discarded, if they can be verified or are within range. If any outliers are dropped from the dataset they will be reported and reasons for their exclusion given.

4.7 ASSUMPTION CHECKING

For the ANCOVA analyses of outcome measurements, the ANCOVA assumptions: homogeneity of regression slopes, homogeneity of variance, normality of residuals and linearity (between the covariate and outcome measure) will be tested. If the assumption of homogeneity of regression slopes has been substantially violated an ANOVA on the difference between the baseline score and follow-up score will be conducted. For substantial violation of the other assumptions an appropriate non-parametric test will be conducted. Note also that if there was a substantial imbalance between the groups for a baseline variable, we would investigate whether the variable was prognostically important, and if so incorporate it into the analysis model.

For the linear regression models, residual plots will be used to determine whether the residuals are normally distributed and that error variance is constant. For the Cox proportional hazards regression for the first fall data, the proportional hazards assumption that the hazard ratio between groups (i.e., PrAISED intervention vs treatment as usual) remains constant over time will be testing by using a log minus log plot.

5. DATA

5.1 DATA COLLECTION AND HANDLING

Data collection and entry onto a MACRO[1] database was undertaken at the individual sites. Cleaning and analysis is undertaken by NWORTH using standard, secure, anonymous procedures for handling participant data. The fully auditable MACRO data management system was used to ensure best practice. The health economics and patients' demographic data will be securely transferred from NWORTH to

[1] https://www.elsevier.com/en-gb/solutions/macro

CHEME (Centre for Health Economics and Medicines Evaluation) for further analysis. For further details, please refer to the Data Management Plan.

5.2 TIME POINTS FOR COLLECTION OF OUTCOME MEASURES

Data was collected at baseline assessment and at a 12 months follow-up for the measures list in tables 1-5.

Table 1: Abilities and Activities measures collected at the baseline assessment and 12 months follow-up assessment

Data / Measures	Refers to	Completed by	Outcome Measure Type (Primary/Secondary)
DAD (Disability Assessment for Dementia)	Participant	Informant	Primary
GAQ (Group Activities Questionnaire)	Participant	Informant	Secondary
LAPAQ (LASA Physical Activity Questionnaire)	Participant	Informant	Secondary
NEADL (Nottingham Extended Activities of Daily Living)	Participant	Participant	Secondary
Pedometer	Participant	Participant	Secondary
TUG and Dual TUG (Timed Up and Go)	Participant	Participant	Secondary

Table 2: Health Status measures collected at the baseline assessment and 12 months follow-up assessment

Data / Measures	**Refers to**	**Completed by**	**Outcome Measure Type (Primary/Secondary)**
Berg Balance Scale	Participant	Participant	Secondary
Blood Pressure	Participant	Participant	Secondary
Frailty	Participant	Informant	Secondary
Medical History and Medications	Participant	Informant	Secondary
Muscle Strength	Participant	Participant	Secondary

Table 3: Psychological and Cognitive measures collected at the baseline assessment and 12 months follow-up assessment

Data / Measures	**Refers to**	**Completed by**	**Outcome Measure Type (Primary/Secondary)**
AES (Apathy Evaluation Scale)	Participant	Informant	Secondary
Personality Big Five Inventory - Short[1]	Participant	Informant	Secondary
CANTAB (Cambridge Neuropsychological Testing Automated Battery)	Participant	Participant	Secondary
CSI (Carer Strain Index)	Informant	Informant	Secondary
FESI – Short (Short Falls Efficacy Scale International)	Participant	Participant	Secondary
HADS (Hospital Anxiety and Depression Scale)	Participant	Participant	Secondary
MoCA (Montreal Cognitive Assessment)	Participant	Participant	Secondary
Verbal Fluency	Participant	Participant	Secondary

[1] This outcome measure was not collected at the 12 months follow up assessment since it is being primarily collect as a potential predictor variable for adherence and due to its nature, it unlikely to change markedly at the 12 months follow up.

Table 4: Quality of Life measures collected at the baseline assessment and 12 months follow-up assessment

Data / Measures	**Refers to**	**Completed by**	**Outcome Measure Type (Primary/Secondary)**
DEMQOL (Dementia Quality of Life)	Participant	Participant	Secondary
DEMQOL – Proxy	Participant	Informant	Secondary
EQ-5D-3L (EuroQol 5 Dimensions scale)[1]	Participant	Participant	Secondary
EQ-5D-5L – For Carer	Informant	Informant	Secondary
EQ-5D-5L – Proxy[2]	Participant	Informant	Secondary

Table 5: Resources related measures collected at the baseline assessment and 12 months follow-up assessment

Data / Measures	**Refers to**	**Completed by**	**Outcome Measure Type (Primary/Secondary)**
CSRI (Client Service Receipt Inventory)	Participant	Informant	Secondary
CSRI (Carer Version)	Informant	Informant	Secondary

[1] The pain item from the EQ-5D-3L will also be analysed separately as a secondary outcome.
[2] The pain item from the EQ-5D-5L Proxy will also be analysed separately as a secondary outcome.

CSRI data (data refers to the participant and completed by the carer), data for the DEMQOL-U Proxy (data refers to the participant and completed by the carer) and EQ-5D-5L Proxy (data refers to the participant and completed by the carer) measures, which will be used for health economic analysis, will be collected at 6 months. The EQ-5D-5L Proxy and the DEMQOL-U Proxy assessed at 6 months will also be analysed at a secondary outcome. Also, for the for health economic analysis, the GAQ (Group Activities Questionnaire) will be administered at the 12 months follow-up only.

Calendar data, which contains information on falls, injuries, amount of exercise engaged in and service use will be collected every month after baseline assessment for 15 months.

5.3 DEFINITIONS AND CALCULATIONS OF OUTCOME MEASURES

For a full definition and respective calculations of each outcome measures please see Appendix 1.

5.4 PARTICIPANT POPULATION

Males and Females, aged 65 or over who meet the following eligibility criteria:

- Have a diagnosis of mild cognitive impairment or Dementia (other than Dementia with Lewy Bodies).
- Are able to communicate sufficiently in English to take part in the research and intervention.
- Are able to walk without human help (using a stick or frame if needed).
- Can see to read.
- Can hear sufficiently to communicate in conversation.
- Can hold and use a pen (i.e., have sufficient dexterity to complete the neuropsychological tests).
- Have a friend / relative / carer willing to take part as an informant, who has contact with the participant for at least one hour per week.
- Are likely to be available for the next year (i.e., do not have plans to relocate or go on a long holiday, or have a life expectancy of less than a year).
- Are free from any comorbidities that would prevent participation (such as severe breathlessness, Parkinson's, severe pain, psychosis).
- Have the capacity to give consent and agreement to participate.

5.5 SAFETY DATA

The PrAISED intervention group has substantially more contact with healthcare staff than the treatment as usual group. This provides more opportunities for ascertaining potential adverse events, resulting in information bias. A direct comparison of AE/SAE rates is therefore difficult to interpret. The number (and proportion) of participants experiencing AE/SAE will be presented for each group. For each participant the timing, severity, expectedness and likelihood that the event was associated with the trial will be reported.

6. STATISTICAL ANALYSES

6.1 ANALYSIS TIME FRAME

Table 6: The expected completion dates for data analysis related tasks

Task	Expected date
Analysis of Internal Pilot	March 2019 – April 2019
Data Extraction and Cleaning for Baseline (excluding Calendar Data)	January 2022 – February 2022
Data Extraction and Cleaning for 12 Months Follow Up (excluding 15 Months Falls Data)	July 2022 – August 2022
Analysis of all Data Excluding 15 Months Falls Data	September 2022 – October 2022
Data Extraction and Cleaning for 15 Months Follow Up Falls and Service Use	October 2022 – November 2022
Analysis of 15 Months Fall Data	November 2022 – December 2022

6.2 BASELINE ANALYSIS

Descriptive statistics of the following variables will be presented for the sample and then split for each of the blinded arms of the trial:

- Gender, relationship status, ethnic group, main language, living alone
- Highest education level, further education/training, time in education

- Comorbid medical diagnoses, medication count
- All the remaining measures listed in Table 1 above

At this stage, we would not expect there to be any difference between the groups so this will be checked to determine if this is the case. Categorical variables will be presented using counts and percentages and continuous variables will be presented with mean, medians, inter-quartile ranges and ranges. No formal statistical testing will take place on baseline data, data will only be used to describe any imbalance.

6.3 INTERNAL PILOT ASSESSMENT

The internal pilot assessment has been described in a separate document: Statistical Assessment Plan for the (PrAISED) internal pilot.

6.4 CONSORT ANALYSIS FLOW DIAGRAM

The CONSORT flow diagram, as shown in section 3.4 above, will be completed. Additionally, the difference in attrition rates between the PrAISED intervention and the treatment as usual groups will be a be analysed using a Fisher's Exact Test.

6.5 ANALYSIS OF THE PRIMARY OUTCOME

The primary outcome is the DAD score, which has been used to measure the participant's disability in activities of daily living from the carer's perspective. To compare the difference between the DAD score at the 12 months follow-up for the treatment as usual and the PrAISED intervention groups, an ANCOVA will be conducted using the stratification variables (i.e., site, co-resident carer & history of falls) and the baseline measure DAD score as covariates. The analysis will be conducted as an intention to treat analysis.

6.5 ANALYSIS OF THE SECONDARY OUTCOMES

To compare the difference between the secondary outcomes listed in Table 1-4 at the 12 months follow-up for the treatment as usual and the PrAISED intervention groups, an ANCOVA will be conducted using the stratification variables (i.e., site, co-resident carer & history of falls), and the baseline measure score as covariates. The above analysis will also be conducted on the single pain items from the EQ-5D-3L (self-rated) and EQ-5D-5L (proxy). All these analyses will be conducted as intention to treat analyses.

In line with Logan et al (2010) and Roberston, Campbell & Herbison (2005) analyses and recommendations, the rate of falls, derived from falls data collected over 12 months (i.e., Month 3 to Month 15 post baseline assessment) will analysed using a negative binominal regression. (Note, that the negative binominal distribution is recommend over the Poisson distribution because the Poisson distribution is over-dispersed.) This will produce an estimate of the incidence rate ratio for group (treatment as usual vs PrAISED intervention) and a confidence interval for the incidence rate ratio estimate.

Like the analysis conducted by Logan et al (2010), time to the first fall data, derived from falls data collected over the over 12 months (i.e., Month 3 to Month 15 post baseline assessment) will be analysed using a Cox proportional hazards regression. This will produce an estimate of the hazard ratio for group (treatment as usual vs. PrAISED intervention) and a confidence interval for the hazard ratio estimate. Both the negative binominal regression of rate of falls and the Cox proportional hazards regression of the time first fall will intention to treat analyses.

6.6 ANALYSES OF SAFETY OUTCOMES

Due to an observed, substantial, reporting bias of safety related events (i.e., participants in the intervention group have greater opportunity to report safety related events) no further in-depth analysis of the safety outcome will be performed. Hence, analysis of safety outcomes will be confined to the descriptive statistics analyses described above in section 5.5. Note also that descriptive statistics for the deaths, hospital admissions, falls and injurious falls and new care home placements with be presented for both arms.

6.7 EXPLORATORY ANALYSIS

An exploratory analyse will be undertaken, if the analysis of the primary outcome yields no statistically significance difference between the treatment as usual and the PrAISED intervention groups. This analysis will involve repeating the analysis of the primary outcome with only the participants from the PrAISED intervention group who have undertook more than 75% of the PrAISED sessions regardless of mode of delivery (i.e., face to face or remote). Similarly, an analysis of the primary outcome with only the participants from the PrAISED intervention group who received more than 75% of intended sessions face-to-face will also be conducted

6.8 SENSITIVITY ANALYSES

In accordance with the advice of Thabane et al (2013), all sensitivity analyses will be conducted on the primary outcome measure. A sensitivity analysis, using a complete cases analysis, will be conducted to assess the effects of missing data imputation on analyses outcome when missing data is greater than 5%. A per-protocol analysis will also be conducted as a sensitivity analysis. To account for survivor bias, a sensitivity analysis, where death will be treated as a DAD score of 0, will be conducted if there is a marked difference in mortality between the two arms. Also, a sensitivity analyses will be conducted by excluding data from the participants who completed 12 months follow up not within the +/- 4 weeks window.

6.9 PROCESS EVALUATION

The process evaluation, adherence and motivation study analyses will be described separately, see Di Lorito et al (2019).

6.10 HEALTH ECONOMICS ANALYSIS

The health economic analysis will be described in a separate document.

7. COVID-19 PANDEMIC RELATED REVISIONS TO THE STATISTICAL ANALYSIS PLAN

This section outlines and describes the revisions to the original statistical plan to accommodate for the COVID-19 pandemic.

7. 1 DATA AND DATA COLLECTION

Additional data have been collected at follow up concerning the effects of the COVID-19 pandemic on the study participant. This COVID-19 related data which concerns infections, social distance and isolation will be analysed descriptively.

Data were also collected at an additional ad-hoc interim follow up timepoint for individuals participating in the trial at the onset of the COVID-19 pandemic, in case further data collection became impossible (between March and September 2020). Participants who were between 10.5 and 12 months through the follow up period, immediately proceeded onto the 12 months follow up and did not complete this ad-hoc interim follow up. The time from baseline assessment to the interim follow up occur varied between participants. Therefore, any analyses using this data will

incorporate the months between the baseline and follow up assessment as a factor in the analysis.

Data collection during the COVID-19 pandemic lockdown period was conducted remotely (via post and telephone) for the ad-hoc interim and 12 months follow up assessments. Face-to- face data collection re-commenced in September 2020 where the site and participants were able and willing, as some sites had clinical and research staff re-deployed to vaccine and other COVID-19 related studies. Data collection was undertaken using personal protective equipment recommended at the time, including masks, gloves, aprons and face shields, and 2 metre social distancing. As a result of collecting the data remotely, data for the following outcome measures were not collected:

- Verbal Fluency
- Blood Pressure
- CANTAB
- Timed Up and Go (TUG)
- Muscle Strength
- Berg Balance
- Pedometer
- MoCA
- Frailty

Follow-up conducted face-to-face under pandemic restrictions did not collect:

- CANTAB
- Pedometer

7. 2 SAMPLE CHARACTERISTICS AND SAMPLE SIZE

In line with the recommendations of Meyer et al (2020), differences in the demographic and baseline characteristics of participants enrolled to the study prior to the COVID-19 pandemic and during the COVID-19 pandemic will be examined using descriptive statistics. Similarly, differences in the demographic and baseline characteristics of participants who withdrew from the study prior to the COVID-19 pandemic and during the COVID-19 pandemic will also be examined.

The lockdown living conditions during the COVID-19 pandemic and the adoption of remote data collection may arguably increase the variability in the primary outcome measure (DAD). The TSC therefore requested, on a TSC meeting held on 02/02/2021, that the standard deviation of the primary outcome measure (DAD) at the 12 months follow up be calculated with the data currently entered into MACRO (N = 204) and the sample size be recalculated using an ANCOVA model. The TSC additionally requested that we calculate the mean difference on the DAD that we had 90% statistical power to detect at that time. The results of these calculations are document in an email to the CI (TSC Requested Calculations.pdf) and did not warrant the original target sample being increased.

7.3 WITHDRAWALS AND MISSING DATA

To investigate whether the COVID-19 pandemic affected withdrawals, the number of withdrawals attributed to the COVID-19 pandemic (e.g., infections) will be compared across the treatment as usual and intervention arms.

To investigate the effects of the COVID19 pandemic on missing data, missing data for outcome measures collected prior to the COVID19 pandemic with missing data for outcome measures collected after the onset if the pandemic. If marked differences exists for a given outcome measures, then a COVID-19 factor (pre-pandemic vs during pandemic) will be incorporated into the modelling and imputation of missing data.

7.4 SUBGROUP ANALYSIS

In line with the recommendations of Meyer et al (2020), a subgroup analysis will be conducted using the follow up data collected prior to the onset COVID-19 pandemic and follow up data collected after the onset of the COVID-19 pandemic. Subgroup analyses will be conducted for all the primary and secondary outcomes that were collected.

7.5 SENSITIVITY ANALYSES

In accordance with the advice of Thabane et al (2013), all sensitivity analyses will be conducted on the primary outcome measure.

Separate sensitivity analyses will be conducted by excluding data from the participants who:

i. Had 12 months follow up data collected before the 12 months follow up (i.e., had their 12 months follow up data collected between 10.5 to 12 months after the baseline assessment).
ii. Had a delayed 12 months follow up data collection due to paused commencement of the intervention due to the pandemic
iii. Self-isolated due to COVID-19

An additional sensitivity analysis will be conducted to model and evaluate the interaction between the mode of data collection at the 12 months follow up (face to face vs remote) and the group allocation (treatment as usual vs intervention).

7.6 EXPLORATORY ANALYSES

Data from the ad-hoc interim follow up where the time from the baseline assessment to follow up varied between participants will be analysed for both the primary and the secondary outcomes. These analyses will additionally incorporate the time between the baseline assessment and the follow up measured to the nearest month as a model factor.

An analysis of the change in primary outcome between the baseline and the 12 months follow up will be conducted that will take into account the ratio of the number of PRAISED intervention sessions that were delivery face to face versus remotely.

7.7 PROCESS EVALUATION

Due to the adoption of remote intervention delivery, the process evaluation will additionally document and descriptively evaluate the impact and effectiveness of therapy delivered remotely (i.e., by telephone or video interview) in relation to therapy delivered face to face.

8. SOFTWARE

All quantitative analysis will be completed using SPSS, Stata and R.

9. REFERENCES

Althouse, A. D. (2016). Adjust for multiple comparisons? It's not that simple. *The Annals of thoracic surgery, 101*(5), 1644-1645.

Bajwa, R. K., Goldberg, S. E., Van der Wardt, V., Burgon, C., Di Lorito, C., Godfrey, M., ... & Harwood, R. H. (2019). A randomised controlled trial of an exercise intervention promoting activity, independence and stability in older adults with mild cognitive impairment and early dementia (PrAISED)- A Protocol*. Trials, 20*(1), 1-11.

Blankevoort, C. G., Van Heuvelen, M. J., Boersma, F., Luning, H., De Jong, J., & Scherder, E. J. (2010). Review of effects of physical activity on strength, balance, mobility and ADL performance in elderly subjects with dementia. *Dementia and geriatric cognitive disorders*, *30*(5), 392-402.

Bono, C.A. , Ried, L.D. , Kimberlin, C.A. , & Vogel, W.B. (2007). Missing data on the Center for Epidemiological Studies Depression Scale: A comparison of four imputation techniques. *Research in Social & Administrative Pharmacy, 3*, 1-27.

Di Lorito, C., Pollock, K., Harwood, R., das Nair, R., Logan, P., Goldberg, S., ... & Van Der Wardt, V. (2019). Protocol for the process evaluation of the promoting activity, independence and stability in early dementia and mild cognitive impairment (PrAISED 2) randomised controlled trial*. Maturitas, 122*, 8-21.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics, 11*(1), 1-21.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.

Li, G., Taljaard, M., Van den Heuvel, E. R., Levine, M. A., Cook, D. J., Wells, G. A., ... & Thabane, L. (2017). An introduction to multiplicity issues in clinical trials: the what, why, when and how. *International journal of epidemiology, 46(*2), 746-755.

Logan, P. A., Coupland, C. A., Gladman, J. R., Sahota, O., Stoner-Hobbs, V., Robertson, K., ... & Avery, A. J. (2010). Community falls prevention for people who call an emergency ambulance after a fall: randomised controlled trial. *BMJ*, *340*, c2102.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, *17*(1), 162.

Meyer, R. D., Ratitch, B., Wolbers, M., Marchenko, O., Quan, H., Li, D., ... & Englert, S. (2020). Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic. *Statistics in Biopharmaceutical Research*, 1-22.

Pitkälä, K. H., Pöysti, M. M., Laakkonen, M. L., Tilvis, R. S., Savikko, N., Kautiainen, H., & Strandberg, T. E. (2013). Effects of the Finnish Alzheimer disease exercise trial (FINALEX): a randomized controlled trial. *JAMA internal medicine*, *173*(10), 894-901.

Rao, A. K., Chou, A., Bursley, B., Smulofsky, J., & Jezequel, J. (2014). Systematic review of the effects of exercise on activities of daily living in people with Alzheimer's disease. *American Journal of Occupational Therapy*, *68*(1), 50-56.

Robertson, M. C., Campbell, A. J., & Herbison, P. (2005). Statistical analysis of efficacy in falls prevention trials. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *60*(4), 530-534.

Russell, D., Hoare, Z. S. J., Whitaker, R. H., Whitaker, C. J., & Russell, I. T. (2011). Generalized method for adaptive randomization in clinical trials. *Statistics in Medicine, 30*(9), 922-934.

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., ... & Debono, V. B. (2013). A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC medical research methodology*, *13*(1), 1-12.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine, 30*(4), 377-399.

10. ACKNOWLEDGEMENTS

This document has been reviewed by members of the PrAISED team.

This paper presents independent research funded by the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research Programme (Reference Number RP-PG-0614-20007). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

APPENDICES

Appendix 1 – Outcome measures summary table

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
MoCA (Montreal Cognitive Assessment)	Assesses different types of cognitive abilities.	Visuospatial and Executive Functioning: 5 points Animal Naming: 3 points Attention: 6 points Language: 3 points Abstraction: 2 points Delayed Recall (Short-term Memory): 5 points Orientation: 6 points Education Level: 1 point is added to the test-taker's score if he or she has 12 years or less of formal education	None	None	26 or higher considered normal	Higher is superior	2
Disability Assessment for Dementia (DAD)	Evaluates the basic and instrumental activities in daily activities of older people with Dementia. Specifically measures daily living	A 40-item measure. Items scoring ranges 0 -1, 0 = No and 1 = Yes. Each subscale is scored by summing items divided by number of applicable items. The Total score is then calculated by summing the three subscale total scores and dividing by applicable items, this score is then converted to a percentage.	3 subscales: Initiation, Planning & Organizing, Effective Performance.	None.	None.	Higher is superior	N/A

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
	tasks in terms of executive functions.	Range of final score 0-100%.					
Nottingham extended ADL scale (NEADL)	Instrumental activities of daily living measure	A 22-item measure. Scoring ranges from 0 – 1, 0 = 'Not at all' OR 'with help' and 1 = 'On my own with difficulty' OR 'On my own'. Total Score is calculated by summing the 22 individual items giving a range of 0 -22. Higher scores represent greater independence.	None.	None.	None.	Higher is superior	4.9
Study Calendar	Study specific measure that monthly captures data on: falls, PrAISED Exercises (completed and duration) and Other Exercises (completed and duration)	N/A	None	None	None		N/A

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
Pedometer	Number of steps (per day)	N/A	None	None	None	Higher is superior	N/A
Timed Up and Go (TUG)	A simple test used to assess a person's mobility and requires both static and dynamic balance.	Time to complete trial in seconds.	None	None	None	Higher is inferior	3.4 secs
Dual Task Timed Up and Go	Test used to assess a person's mobility and requires both static and dynamic balance while performing a task (either counting backward or naming words beginning with a given letter)	Time to complete trial in seconds.	None	None	None	Higher is inferior	N/A

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
SHARE Frailty Index	A measure of frailty	5 items assessing: Exhaustion, weight loss, weakness, slowness and low activity Specific scoring algorithm used to calculate the index	None	None	None	Higher is inferior	N/A
Berg Balance Test	Used to measure balance among older people with impairment in balance function by assessing the performance of functional tasks.	A 14-item measure. Each item is scored on a five-point scale, ranging from 0-4. "0" indicates lowest level of function and "4" indicates the highest level of function. Final value calculated by summing individual scoring components. Range of final score 0 – 56.	None.	None.	41 - 56 = low fall risk 21 - 40 = medium fall risk 0 - 20 = high fall risk	Higher is superior	4 - 7
Short Falls Efficacy Scale International (Short FES-I)	Used to measure the level of concern about falling during social and physical	A 7-item measure. Items are scored on a four-point Likert scale ranging from 1 – 4. 1 = not at all concerned and 4 = very concerned.	None.	None.	7-9: Low concern 9-13: Moderate concern 14-28: High Concern	Higher is inferior	N/A

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
	activities inside and outside the home whether or not the person actually does the activity.	Final score calculated by summing the 7 individual scoring components. Range of final score 7 – 28, higher scores being a greater fear of falling.					
EQ-5D-3L	Used to measure health state of participant.	5-digit outcome with one value representing each of the 5 domains. A utility value can also be calculated to give a single value with a maximum of 1 representing full health. Visual analogue scale represents patients perceived health from 0 (worst) to 100 (best) 3 levels; no problem, some problems & extreme problems	5 subscales: mobility, self-care, usual activity, pain & discomfort and anxiety & depression.	None.	None.	Higher is superior	0.074

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
EQ-5D-5L	Used to measure health state of participant.	5-digit outcome with one value representing each of the 5 domains. A utility value can be calculated to give a single value with a maximum of 1 representing full health. Visual analogue scale represents patients perceived health from 0 (worst) to 100 (best) 5 levels; no problems, slight problems, moderate problems, severe problems and unable	5 subscales: mobility, self-care, usual activity, pain & discomfort and anxiety & depression.	None.	None.	Higher is superior	0.037
DEMQoL	Measures health-related quality of life for people with Dementia.	A 28-item measure. Items are scored on a 4-point scale ranging from 1 – 4: 1 = 'Not at all' and 4 = 'A lot'. An additional Item (29) has been added and is scored on a 1 – 4 scale with 1 = 'very good' and 4 = 'poor'. Items 1, 3, 5, 6, 10 and 29 are reversed scores. To calculate the total score the individual items are summed together, so the total score can	None.	If at least 50% of items are complete, impute with person-specific mean of completed items	None.	Higher is superior	5 - 6 points

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
		range from 1 – 116. The higher the score the better health related quality of life.					
DEMQoL Proxy	Measures health-related quality of life for people with Dementia.	A 31-item measure. Items are scored on a 4-point scale ranging from 1 – 4: 1 = 'Not at all' and 4 = 'A lot'. Item 31 is scored on a 1 – 4 scale with 1 = 'very good' and 4 = 'poor'. Items 1, 4, 6, 8, 11 and 31 are reversed scores. To calculate the total score the individual items are summed together, so the total score can range from 1 – 124. The higher the score the better health related quality of life.	None.	If at least 50% of items are complete, impute with person-specific mean of completed items	None.	Higher is superior	N/A
DEMQOoL-U Proxy	Health state classification and utility score for people with dementia	The health state is derived from combining the scores from 4 questions to derive a health state. The questions concern: liveliness, forgetfulness, appearance and frustration. The questions are scored from 1 to 4. The utility score can be calculated to give a single value	None	None	None	N/A	0.02 - 0.05

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
		with a maximum of 0.937 representing full health.					
CANTAB Paired Associated Learning (PAL)	A neuro-psychological assessment	Paired Associated Learning (PAL) total errors (scoring range 0-70) PAL first attempt memory score (scoring range 0-20) PAL number of patterns reached (scoring 2-8)	None None None	None None None	None None None	Higher is inferior Higher is superior Higher is superior	N/A
CANTAB Spatial Span Test (SST)	A neuro-psychological assessment	Spatial Span Test (SST) Forward Span Length (score range 3-9). The longest sequence successfully recalled by the subject. Forwards variant only.	None	None	None	Higher is superior	N/A
CANTAB Multi-Tasking Test (MTT)	A neuro-psychological assessment	Multi-Tasking Test (MTT) Right Left Task Median switching block (in milliseconds) Multi-Tasking Test (MTT) Right Left Task Median congruent (in milliseconds)	None None None	None None None	None None None	Higher is inferior Higher is inferior Higher is inferior	N/A

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
		Multi-Tasking Test (MTT) Right Left Task Median incongruent (in milliseconds)					
Client Service Receipt Inventory (CSRI)	Self-reported service user data to evaluate and cost service use.	Collection of this data is essential for HE analysis but will not be subjected to any "scoring"	N/A	N/A	N/A	N/A	N/A
Hospital Anxiety and Depression Scale (HADS)	Used to detect states of depression, anxiety and emotional distress amongst patients.	Two subscales of 7 items measured on a 4-point scale ranging from 0 – 3. Scores for the subscales are calculated by summing items within scale giving a range of 0 – 21 for the sub scales. Items 7 and 10 are reversed scored. Total score calculated by summing the two subscales giving a total range of 0 – 42.	2 sub scales: Depression and Anxiety. Depression subscales = 1, 3, 5, 7, 9, 10, 13. Anxiety subscales = 2, 4, 6, 8, 11, 12, 14.	A single missing item from a subscale may be replaced by the mean of the remaining six items. If more than one is missing the subscale is invalid.	Threshold for *subscales;* 0-7 = Normal 8–10 = Borderline abnormal 11-21 = Abnormal (case)	Higher is inferior	N/A

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
Carer Strain Index (CSI)	Tool that measures strain related to care provision	13 item questionnaire. Sum of yes answers. Total core range from 0 to 13.	None	None	7 or higher indicates a high level of stress	Higher is inferior	N/A
Apathy Evaluation Scale (AES)	Developed to provide global measures of apathy in adults and elderly individuals	18 item questionnaire. Items scored 1 to 4. Total score is the sum of scores, which can range from 18 to 72	None	None	None	Higher is superior	N/A
Big Five Inventory – Short version (BFI-10)	Inventory designed to measure the Big Five dimensions of personality	5 subscales of 2 items measured on 5 point scale range from 1 – 5. Scores for the subscales are calculated by summing items.	5 sub scales: Extraversion Agreeableness Conscientiousness Neuroticism Openness to Experience	None	None	N/A	N/A
LASA (Longitudinal Aging Study Amsterdam) Physical Activity Questionnaire (LAPAQ)	Developed to measure physical activity	Sum of 6 physical activity (in minutes) over a 2 week period. The 6 activities are: walking, cycling, gardening, light and heavy household work and sports.	None	None	None	Higher is superior	N/A

Outcome measure	Definition	Scoring	Subscales	Missing value rules	Thresholds	Interpretation	Minimal Clinically Important Difference
Verbal Fluency	Assessment of Verbal Fluency	Number of correctly named animals (within 1 minute)	None	None	None	Higher is superior	N/A
Group Activities Questionnaire (GAQ)	Study specific measure that assesses individual's engagement in group activities	N/A	None	None	None	N/A	N/A