OPERA

Statistical Analysis Plan

Approved Version 1.1: 15/09/11
Version 1.0: 11/07/2011

Approved by:
Signature: _____ Date: ___/___/_____
Approved by:
Signature : _____ Date: ___/___/_____

Trial Statisticians: Stephen Bremner, Sandra Eldridge, Karla Diaz-Ordaz

This document was created based on the Mental Health & Neuroscience Clinical Trials Unit (MH&N CTU) Analysis Strategy template (version 1.5; 13/02/2008) based on an earlier version developed by Rebecca Walwyn and Sally Lee.

A randomised trial of an exercise intervention for older people in residential and nursing accommodation

Contents

1 INTRODUCTION

1.1 Purpose of statistical analysis plan

The purpose of this document is to provide details of the statistical analyses and presentation of results to be reported within the principal paper(s) of the OPERA trial. Subsequent papers of a more exploratory nature (including those involving baseline data only) will not be bound by this strategy but will be expected to follow the broad principles laid down in it. Any exploratory, post-hoc or unplanned analyses will be clearly identified in the respective study analysis report.

The structure and content of this document provides sufficient detail to meet the requirements identified by the International Conference on Harmonisation (ICH) and the PCTU SOP (PCTU/07).

The following were reviewed in preparation for writing this document:
Trial application submitted *15th March 2006*
ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials
ICH E9 Guidance on statistical principles for clinical trials
ICH E3 Structure and content of clinical study reports
CONSORT guidelines for the reporting of randomised trials, including extensions to cluster-randomised trials and trials of non-pharmacological treatments

1.2 Members of the writing committee

Stephen Bremner, Karla Diaz-Ordaz, Sandra Eldridge and Martin Underwood were primarily responsible for writing the Statistical Analysis Strategy with Stephen Bremner responsible for writing the computer code implementing the analysis strategy, and implementing the strategy at the point of analysis. A senior statistician (Richard Hooper) within the Pragmatic Clinical Trials Unit (PCTU) also reviewed and signed off the analysis strategy prior to a final check by the TSC statistician, Sally Kerry. If decisions are required during the course of the analysis they will be discussed with a statistician within the PCTU, independent of the trial.

This document has been developed prior to examination of trial data and will not be implemented prior to final approval. Much of it is derived from the research protocol (Underwood, 2010).

1.3 Summary

Exercise is a promising non-medical approach to the management of depression. Plausible mechanisms for its possible effect include improved social contact, a diversion from negative thoughts, and the physiological effects on neurotransmitters such as monoamines and endorphins. In this trial we are testing a pragmatic intervention, reflective of current best practice, consisting of training for residential or nursing home (RNH) staff to support the building of safe physical activity into the RNHs' normal routine; and a twice-weekly formal exercise class led by a specially trained physiotherapist.

This is a cluster-randomised trial, with the RNH as the unit of randomisation and residents as the unit of assessment, to study the impact of a whole RNH intervention to increase exercise on the prevalence of depression and the remission of existing depression.

1.4 Changes from planned analysis in the protocol

We will use mixed models rather than generalised estimating equations (GEEs) for the analysis. One motivation for this is the less strict assumption about missing data made (i.e. data missing at random rather than missing completely at random).
One of the co-primary outcomes was changed to a secondary outcome with approval of the TSC [See TSC minutes from 22/09/2009]. Following an early decision [See TMG minutes from 01/04/2009] part-way through the trial to exclude dementia homes, only one dementia home was included.

There was therefore only one dementia-specialist home in the study, and though strictly speaking, dementia-specialist status was included in the minimisation algorithm, it effectively had no influence in the randomisation schedule (as there was no adjustment done on "dementia specialist" status homes between the two arms). As such, we need not adjust for it.
Exclusion criteria for homes were also modified from protocol (minutes 11/02/2009) though this should not affect planned analyses.

2 STUDY OBJECTIVES AND ENDPOINTS

2.1 Study objectives

2.1.1 Primary objectives

Our primary objective is to compare depression levels between intervention and control homes. See Section 2.2.1 for the definition of depression used in this study. Three comparisons will be made:

(a) Compare the prevalence of depression in intervention homes with that in control homes in all residents contributing data twelve months after homes were randomised and who had also been resident by time of the home's nine month assessment **(cross-sectional comparison)**.
(b) Compare the change in number of depressive symptoms at 6 months in residents who were depressed at baseline **(cohort comparison)**.
(c) Compare the change in number of depressive symptoms in residents who were present at baseline and at twelve months after randomisation. **(cohort comparison)**

We distinguish between the **cross-sectional analysis,** required for objective (a), and **cohort analysis** required for the objectives (b) and (c).

For the **cross-sectional analysis,** outcome data have been collected on the majority of residents in the homes at 12 months. Those residents who had entered a home after a pre-specified cut-off date have been excluded on the grounds that they would not have experienced a sufficient dose of the intervention. This cut-off date was around three months prior to the outcome data collection in most homes but between two and three months prior to outcome data collection in seven homes in which final outcome data collection occurred during the 12th month rather than at the end of the 12th month.

In the **cohort analyses**, a different, but overlapping set of residents (those who had baseline data collected pre-randomisation) provide baseline data. In the cohort analysis those who provide outcome data are the same individuals who provide baseline data prior to randomisation.

This distinction, between cross-sectional and cohort analyses, also applies to secondary outcome analysis and we describe analyses more fully later in this document.

2.1.2 Secondary objectives

To compare the following between intervention and control arms:

- remission of depression by 6 months,
- number of depressive symptoms at 12 months

- cognitive function (at 6 and 12 months),
- health-related quality of life (at 6 and 12 months),
- mobility and exercise tolerance (at 12 months),
- chronic pain (at 6 and 12 months),
- fear of falling (at 6 and 12 months),
- social engagement (at 6 and 12 months).

For all residents with consent/assent to examine the medical and care home records:
Present descriptive statistics on the prescribing of antidepressants
Present descriptive statistics by arm on hospital admissions.
Apart from remission of depression, the primary endpoint will be 12 months for each of these secondary outcomes.

2.1.3 Safety Objectives

To compare the following between intervention and control arms:

(a) Summarise incidence of injurious falls as indicated by peripheral fractures, summarised by home
(b) Compare mortality rates in intervention and control homes.

2.1.4 Exploratory objectives

Whilst listed here, these will not be dealt with in this analysis plan and will be considered in a separate analysis plan to be written after the main analysis outlined in this plan has been completed.
To assess the effect of dose of exercise intervention classes on outcomes.
To assess whether social engagement and prescribed antidepressants (DDD) mediate the effect of exercise on depression and cognition.

2.2 Outcome measures

The following table (table 1) details the primary outcome measure (GDS) and various secondary outcome measures, their data types, and at which end-points they apply.

Table 1: outcomes, end-points and data types

	Outcome measure	At 12 months in all present at end of study	Change at 6 and 12 months in all present at baseline	Change at 6 months in those depressed and present at baseline	Remission of depression at 6m in those depressed at baseline	All residents who provided consent at any time
primary	GDS	dichotomous	continuous (12 months)	continuous		
secondary	GDS	continuous			dichotomous	
	MMSE	continuous	continuous			
	EQ-5D *	continuous	continuous			
	SPPB	continuous	continuous**			
	Pain rating today	ordinal (5 points)	ordinal (5 points)			
	Social engagement	ordinal (6 points)	ordinal (6 points)			
	Fear of falling	dichotomous	dichotomous			
safety	Peripheral fractures					rate
	All-cause mortality					rate

* both proxy and self-completed
** at 12 months only

2.2.1 Primary outcomes

The primary outcome measure is the Geriatric Depression Scale 15 (GDS-15). This brief scale/score consists of 15 simple yes/no questions and has been well

validated in residential situations. It avoids using potentially somatic features of depression which may be misleading in this age group, focusing more on mood and functional symptoms of depression.

The GDS-15 can be interpreted as an indication of the presence/absence of depressive mood, though this is not a substitute for proper clinical diagnosis. A score of five or above appears to give the best sensitivity and specificity.

- Binary outcome: For the purpose of these analyses, we class a resident as 'depressed' if they have a GDS-15 score of five or above when answering at least 13 items. For those who answered less than 13 items but at least 10, we apply the following rules:
  "Depressed"=Yes if GDS≥4 if they answered 11 or 12 items.
  "Depressed"=Yes if GDS≥3 if they answered 10 items.
  GDS-15 scores are not valid if less than 10 items are answered, in which case they are set to missing.

- Continuous outcome: GDS-15 for the second and third co-primary outcomes
  2nd: change in depressive symptoms* at 6 months in those residents depressed at baseline.
  3rd: change in depressive symptoms* at 12 months in those residents present at baseline.
  * Each item is a depressive symptom (either present (scores 1) or absent (scores 0)).
  If the number of items completed, n is such that 9<n<15, the total score is to be rescaled i.e. GDS' = 15*GDS/n

2.2.2 Secondary outcomes

All of these have both 12-month (primary) (except remission of depression) and 6-month endpoints.

- Remission of depression at 6 months in those depressed at baseline (measured using GDS-15).
- Depressive symptoms at 12 months (measured using GDS-15)
- Cognitive function: measured using the Mini Mental State Examination (MMSE) (treated as a continuous variable). *NB if 15 or more items are missing, the total score should be set to missing.*
- Mobility: the effect of the programme on mobility is assessed using the Short Physical Performance Battery (SPPB) as a continuous variable.
- Falls:
  a) Fear of falling by asking participants "Are you afraid of falling?" requiring only a simple yes/no response from participants.
  b) Rate of peripheral fractures as a marker for injurious falls identified from both RNH (aggregated at the home level) and Primary Care Trust records (at patient-level).
- Pain: Pain today, on a five-point numerical rating scale.
- Health-related quality of life as measured by (a) self-reported EQ-5D (b) proxy EQ-5D

- Social engagement, measured by the Social Engagement Scale designed for use in nursing and residential home care gives an indication of the involvement of a resident in activities within a nursing or residential home. The Social Engagement Scale uses six items from the Minimum Data Set residential assessment instrument (MDS-RAI).

  The following outcomes will be described but will not be subject to formal statistical modelling.

- Medication use: regular medications data collected from RNH records at baseline, three, six, and nine months after randomisation/study entry, and at the end of the study. For anti-depressants, other psychoactive drugs, NSAIDs and other analgesics, we will convert these into the number of defined daily doses (DDD) used over one year (http://www.whocc.no).

- Hospital admissions: we will extract data on cause and duration of any hospital admissions during the study period from participants' hospital records. We will code these admissions into Diagnosis Related Groups (DRGs) or Health Resource Groups (HRGs) as appropriate to identify any fractures, and for the economic analysis (see separate health economics analysis plan).

2.2.3 Safety outcomes

Deaths

Records of all those from whom we obtained consent/assent to examine were flagged at the Medical Research Information Service (MRIS), and assessed in order to identify any differences in mortality between the two groups. For those who died in their RNH we asked the home for a brief description of how they died; for those who died in hospital we extracted this information from their hospital records. Medical members of the study team, blind to participants' allocation, assessed these reports. In the event that any death was deemed to be exercise-related on the basis of these brief reports, we made a detailed assessment as to whether it was related to our programme.

Peripheral fractures

Data on these were collected from two sources (i) aggregated at home level and events counted in all residents, not only those who consented/assented to their data being used. (ii) at the individual level in all residents who provided consent/assent at any time for their medical records (held by the PCT) to be accessed.

3 STUDY METHODS

3.1 Overall study design and plan

Target for randomisation:	77 RNHs (78 actual)
Date of first randomisation:	19th February 2009
Date of last randomisation:	30th April 2010
Trial design:	cluster randomised, parallel group
Blinding:	Clinical researchers were blinded to allocation to interpret records on deaths.

Randomised Interventions: exercise programme + depression awareness programme vs. control intervention (Depression awareness programme)
Planned allocation ratio: 2 intervention homes : 3 control homes

3.2 Selection of study population

3.2.1. Selection of individuals
Inclusion criteria for residents:

- Permanent resident in RNH
- Aged 65 or over
- Consent/assent to be assessed
- Consent or assent to participate in baseline assessment or to provide medical record data
- We are including non-depressed residents in the exercise programme

Exclusion criteria for residents:

- Problems communicating by any means
- Non-English speakers for whom a translator is not available
- Terminal or other serious illness
- Those with a very limited life expectancy

3.2.2. Selection of clusters (i.e. homes):
All RNHs in outer NE London (Barking & Dagenham, Havering & Redbridge and Waltham Forest), NHS Coventry, NHS Warwickshire, and Coventry & Warwickshire Partnership Trust were invited to participate. There are over 80 RNHs in Barking & Dagenham and neighbouring Havering. There are around 135 RNHs in Coventry & Warwickshire Partnership Trust's locality. One home in Essex was also included.

Inclusion criteria:

- Initially homes with more than 15 beds, but 70 or fewer beds; during the study (TMG minutes 01/04/2009) this was revised to 20-60 grounded at the lower end by our recruitment experience and at the upper limit by practicalities of delivering the intervention in large homes.

Exclusion criteria:

- Homes catering mostly to non English-speaking residents.
- Any homes with fewer than 16 beds (25 homes, of which none are dementia homes).
- Any dementia homes with fewer than 50 beds. However, given the time and cost involved in assessing dementia-specialist home residents, and delivering the intervention, only one such home was recruited.

3.3 Method of treatment assignment and randomisation

Randomisation was stratified on centre (i.e. London or Warwick) and RNHs (clusters) were allocated using minimisation with a random element (70% probability that home will be allocated to group that minimises the imbalance. We minimised by size and type of home (local authority/voluntary/private & care home/private & nursing home)

Table 2 Design variables

factor	Number of categories	Details of Categories
Centre	2	London or Warwick
Home type	4	Local authority, voluntary, private & care home, private & nursing
Size	2	≤32, >32 beds

The randomisation was carried out by an independent statistician at the PCTU using the Minim programme written by Stephen Evans, Simon Day and Patrick Royston available from http://www-users.york.ac.uk/~mb55/guide/minim.htm.
Research nurses provided details of a home to be randomised by email directly to the independent statistician. Once the home was randomised, the research nurse was notified by e-mail within a week of their request.

3.4 Treatment masking (blinding)

All baseline assessments for the cohort comparison were collected blind to treatment allocation. Our primary measure, GDS-15, and other questionnaire data at follow-up were collected by research nurses not blinded to the RNH's allocation status. Our interpretation of cause of death data was blind to allocation.

For new residents joining the study after randomisation allocation concealment was impossible because all RNH staff and study staff visiting the RNHs were aware of the home's allocation. We tried to protect against recruitment bias resulting from lack of allocation concealment by ensuring that we were aware of all new residents and reasons for exclusion from the study were monitored.

We will compare, descriptively or by graphical methods, the GDS-15, MMSE and SPPB scores of residents recruited into the study post randomisation with those recruited pre-randomisation, by arms, for evidence that (a) residents recruited post-randomisation differ between arms and (b) the ratio of participants assenting to those consenting changes over time. We hypothesise that less effort may be made to recruit residents at 9 months than at 6 months, than at 3 months, than pre-baseline. Assent requires more effort to attain than consent and so this ratio may decrease over time.

3.5 Sample size determination

Because few RNH residents were thought to move out of residential accommodation we anticipated good follow-up rates. This population has a high mortality, up to 34% per year (Rothera, 2002); additionally, for some, their health will deteriorate so that they are no longer able to complete some, if not all, of the follow-up assessments.
All of our sample size estimates included an inflation factor to account for clustering effects of residents within RNHs.
A conservative value of 0.05 for the ICCs for the different outcomes was used. The inflation factor also depends on the average cluster size and the variation in cluster size. Our average cluster size is different for our three outcomes relevant for the original sample size calculations outlined in table 3.

Table 3 Sample size calculations in the original protocol

	A) To show a reduction in the proportion of participating residents depressed (GDS-15 ≥ 5) at the end of the study from 40% to 25%	B) To show an increase in the remission rate after six months from 25% to 40% in those depressed at baseline	C). To show a mean reduction in GDS-15 score of 1.2 after twelve months in those depressed at baseline
Power	80%	80%	80%
Significance	5%	5%	5%
Simple sample size	343	343	280
Mean cluster size at follow-up	15.0	5.4	4.5
Inflation factor	1.7	1.22	1.175
Total number required at follow-up	583 with complete assessments	418 with depression at baseline & complete assessments	330 with depression at baseline & complete assessments
RNHs required	39	77	74

An interim sample size calculation revealed that we had a smaller than anticipated cluster size and in order to be sufficiently powered for outcome B, we would require an additional 24 RNHs (see Table 4). The HTA was not willing to fund an extension. With the agreement of the TSC, we changed outcome B to be a secondary outcome. We also added outcome C (table 4) which is similar to outcome C (table 3), but measured at 6 months rather than 12 months, to compensate for the smaller than expected cluster size, and the greater loss to follow-up than anticipated. Note, outcome C in table 3 becomes outcome D in table 4.

Table 4. Revised sample size calculations

ICC = 0.053	A) To show a reduction in the proportion of participating residents depressed (GDS-15≥5) at 12 months from 46.1% to 28.8%	B) To show an increase in the remission rate after six months from 25% to 40% in those depressed at baseline	C) To show a mean reduction in GDS-15 score of 1.2 after 6 months in those depressed at baseline: from 7.27 to 6.07 (SD=2.21)	D) To show a mean reduction in GDS-15 score of 1.2 after 12 months in all participating residents: from 4.5 to 3.3 (SD = 3.12)
Power	80%	80%	80%	80%
Significance	5%	5%	5%	5%
Simple sample size	281	342	112	224
Anticipated loss to follow-up	10%, 30%	20%	20%	30%
Mean cluster size at follow-up	9.6, 7.5	3.9	3.9	7.5
Inflation factor	1.46, 1.34	1.16	1.16	1.34
Total number required at follow-up	409 (378) with complete assessments	396 with depression at baseline & complete assessments	130	301
RNHs required	43	101	33	41

4 DATA COLLECTION

4.1 Baseline

The following measures are assessed by asking the resident directly: GDS-15, EQ5D, MMSE, fear of falling, current pain, and a brief physical assessment (SPPB).

The following measures are assessed by asking the carer:
Barthel Index, (proxy) EQ-5D, Social Engagement Scale

We collected demographic data (age, sex, ethnicity, age left school (proxy for social class)) and data on length of residence, fee status and current medication from the RNH records for all those for whom we have consent/assent to access their records.

4.2 Follow-up

See the individual case report forms for a full list of all variables which were measured.
The complete list of case report forms is as follows (grey titles irrelevant for this analysis plan):

Home Data Collection
Randomisation Record
Next of Kin Expression of Interest Form
Initial Assessment
Consent Form (Resident)
Assent Form
Assent Form (Routine Data Only)
Baseline Data Collection Form 1 (Medical Records)
Baseline Data Collection Form 2 (Carers)
Event Notification
Study Register
Follow-Up Data Collection (6 months)
Follow-Up Data Collection (12 months) [Different format to the other follow-up forms (contains a field for the resident's NHS number)]
Follow-Up Data Collection Form 2 (Carers) (at 3, 6, 9 and 12 months)
Follow-Up Record Examination Data Collection
Physical Activity Assessment
Assessment Register
Group Profile (Baseline)
Group Profile (Handover)
Recommendation Form
Supporting Information for OPERA Research Physiotherapist's referral recommendations
Group Attendance Register
Clinical Record Form
Depression Awareness Training Follow-Up
Carer Training Attendance Sheet
Depression Awareness and Activity Training Follow-Up
Depression Awareness Training Feedback (same form for intervention and control)
Physiotherapist Recording Form
Depression Score Notification

4.3 Timing of data collection

Residents, recruited between January 2009 and May 2010, underwent baseline assessments by the recruitment team to establish eligibility and once all had consented/assented/refused, the whole home was randomised to either the exercise intervention or depression awareness training. Follow-up visits were made to all homes at 3, 6, 9 and 12 months post-randomisation. Repeated attempts within one month of the assessment date were made as necessary by the recruitment team to obtain data from participants who were unavailable for assessment at the fixed follow-up points (i.e. at 3, 6, 9 or 12 months). At these fixed follow-up points (except 12 months), residents new to the home since the last visit were invited to participate, baseline assessments were performed and consent/assent was sought. NB this does not include existing study participants who had moved into the home during this time. Furthermore, residents that had previously refused consent/assent could change their mind and become study participants subject to a satisfactory baseline assessment and consent/assent being given.

5 GENERAL ISSUES FOR STATISTICAL ANALYSIS

All analyses will be conducted two-sided and significance interpreted at the 5% level.

5.1 Blinding of the statistical analysis

The study statistician (SB) was not blinded in this study because of (a) the allocation ratio making obvious which group was which and (b) SB handling data to produce the DMC reports.

5.2 Analysis populations

We are conducting three distinct types of statistical analyses within this trial: **cross-sectional** analyses, **cohort** analyses and analyses on **safety** outcomes. Here we define the populations pertaining to each. We define an Intention to Treat Population (ITT) although no strict ITT analysis is planned as part of the main analysis of this trial. We do not describe a per protocol population – no per protocol analyses are planned.

5.2.1 Intent-to-treat analysis population

For each effectiveness outcome, populations for the cross-sectional and cohort analyses are defined as follows:

(a) Cross-sectional analyses –
A strict intention to treat population is difficult to define in a cluster randomised trial, as our intention was to treat all those resident in the home, but we can only include those who agree to participate. For the **cross-sectional analysis,** outcome data have been collected on the majority of residents in the homes at 12 months. Those residents who had entered a home after a pre-specified cut-off date i.e. after 9 months post-randomisation, have been excluded on the grounds that they would not have experienced a sufficient dose of the intervention. This cut-off date was around three months prior to the final outcome data collection in most homes but between two and three months prior to outcome data collection in seven homes in which final outcome data collection occurred during the 12th month rather than at the end of the 12th month. The intention-to-treat population comprises all those resident in the home who consented to intervention (or data collection) at or prior to the cut-off assessment date at their home. Residents that moved will be considered in the home that they have moved to providing that they have been resident there prior to the cut-off date. If they were not resident in the home prior to the cut-off date they will be analysed as if they were still resident in the home in which they originally gave consent.

(b) Cohort analyses –
The intention to treat population is those who provided data at baseline. For those residents that move, they will be assigned to the home in which they started out. For most outcomes the ITT population is all participants who provided the outcome measure of interest prior to randomisation but for the outcome 'remission of depression' it is residents depressed at baseline.

5.2.2 Available-case analysis population

By the nature of the participants included in the study, many will not be able to provide complete data at each time point. For each outcome the available case population will be defined by the number of people able to provide data for that outcome, thus the population will be different for each outcome.

(a) Cross sectional analyses-
Therefore, for this type of outcome, we use intention to treat principles, and include all those residents in the home who consented to intervention (or data collection) at or prior to the 9 month assessment date at their respective home and who have a valid outcome measure at the end of the study (12 months) (i.e. complete case following the ITT principle). Residents who switch homes will be included in the analysis with the home in which they move to if they have been resident in this home prior to the 9 month cut off. Otherwise they will be analysed as if they were still in the home in which they were originally randomised as long as there is an outcome measure collected within 2 months either side of the point 12 months from randomisation.

(b) For cohort analyses-
We will include all those who gave consent/assent, with a valid baseline measure of outcome before randomisation, resident in a participating home and with a valid outcome measure at the relevant time point (either 6m or 12m). Residents who move to another trial home and are therefore still followed up will be analysed according to their original allocation (i.e. either intervention or control, ITT principle) in the home as if they were in the home in which they first resided. For remission of depression this includes only those who had symptoms of depression at baseline.

5.2.3 Per protocol analysis population

We have not defined a per protocol population; it is unclear what such a population would be.

5.2.4 Safety population

For the safety analyses the population is defined as all those in the home at any time during the intervention period with exposure given by the length of time within the home during the intervention period i.e. person months of exposure, which will form the denominator

The safety outcomes are deaths and peripheral fractures, both classified as serious adverse events (SAEs). For deaths the study population is all residents present in the home during the intervention period.
Data on peripheral fractures will be collected in two ways and different populations are defined for each

a. Events in study participants (data available at the level of the individual)
b. Events in all residents (data available aggregated at the level of home).

For deaths and for peripheral fractures where the event is measured on all participants the denominator is person years of exposure. This allows us to include residents who were neither present at the beginning nor the end of the study.

5.2.5 Other populations

Unconventional movements of some residents through the trial need to be accommodated in the analysis.

(i) All residents in one home got moved temporarily to another home in the opposite trial arm and were moved back to their original home before the end of follow-up. They will be analysed in the home where they resided at the outset.

(ii) Residents who move to another study home part way through will have two follow-up reference points. E.g. study participant X had been living in home A since pre-randomisation. In month 8, X moves to home B (which may or may not be in the same trial arm as home A) which had been randomised only 5 months previously. Therefore, when home B is due for its 6-month assessment visits, resident X will be due a 9 month assessment. Follow-up (12 months) would be due to end for participant X at the 9 month assessment of home B. However, for the cross-sectional analysis, assuming resident X is still present, they may have additional follow-ups and their latest follow-up will be the one that is included in the 12 month assessment of home B. Resident X, in the cohort analysis, would be analysed as belonging to home A, not B. In the cross-sectional analysis, X is treated as residing in B unless they have been in B for a relatively short time (see sections 5.2.2 and 5.2.3).

5.3 Database

5.3.1 Description

Questionnaires were completed on paper with the resident. Data were entered directly onto a study laptop after the assessment time with the resident by the recruitment team.
The database is stored in Microsoft Access 2007 as a series of separate tables.

5.3.2 Data quality

Quality assurance checks will be undertaken by Warwick CTU to ensure the integrity of randomisation, data entry procedures and data collection.

A 10% random sample will be checked against paper records to identify data entry errors. At the level of the home, the records of 8 homes will be checked. Should the error rate exceed 0.1% of all data items recorded, a further sample of 8 homes will be checked.
At the level of the resident, a 10% random sample of residents' records will be selected across all homes. Data entered on the GDS-15 will be checked at item level. Should the error rate exceed 0.1%, a further 10% sample will be drawn for checking.
For the other outcome measures and covariates, we will also tolerate an error rate of 0.1%.

We will fit linear regression models of the primary outcome measure (GDS-15) and main secondary outcome measure, MMSE, to see if there is any association with data input clerk (coded using dummy variables). This could reveal to us if any clerks have a tendency to input consistently high or low values for these outcomes or if there is a preponderance of missing values or zeros.

5.3.3 Database freeze

The statistician responsible for the analysis will conduct or oversee additional data checks. These include range checks, logical and consistency checks which may not be

picked up by checks performed at the individual patient level by the research staff that collect and enter the data.

Once all 12 month follow-up data have been checked and errors corrected, the database will be frozen in compliance with PCTU SOP 09c (data freeze, handover and lock).

5.4 Analysis software

The analysis will be carried out using mainly Stata version 10.1 or 11 (StataCorp. 2009). However, some analyses may be carried out using MLwiN, a specialist programme for handling multilevel data.

5.5 Methods for withdrawals, loss to follow-up and missing data

Numbers lost to follow up and withdrawal from the study will be described by arm to try and establish if there is difference in attrition rates.
Cross-sectional: N/A
Cohort analyses: as all cohort analyses will be modelled using maximum likelihood methods, outcomes will not be imputed as we do not, at this stage, wish to assume that measurements are missing at random (MAR).

5.6 Method for handling centre effects

The two recruiting centres (London and Warwickshire) will be distinguished in the analysis by a dummy variable.

5.7 Method for handling randomisation stratification or minimisation factors

All analyses will include the minimisation factors and the location stratification variable (i.e. centre (see 5.6)), which are categorical and defined at cluster level. Dementia specialist care home status (Yes/No) will not be included as there was, by design, only one such home.

5.8 Method for handling clustering effects

We will use mixed effects models with a random effect for RNH.

It is possible that there may be additional clustering effects in intervention homes only, because the exercise classes are group activities and there may be therapist effects. However, exactly who delivered each class in not recorded, though each home's activities are overseen by a primary therapist who has the most influence on how the classes are run. It is not possible to disentangle these effects, if they exist, as physiotherapists on annual or sick leave were replaced temporarily by other physiotherapists.
Therefore, we will simply quantify and describe the clustering for each outcome overall by study arm and present 95% confidence intervals for these to check if there is substantially greater clustering in the intervention arm than in the control arm.

5.9 Method for selecting other variables that will be adjusted for

The variables to be included in the model for each outcome are specified in table 5.

Variables that were considered by the trial team, *a priori*, to be related to the outcome will be included. The most important one (for the cohort analyses) is likely to be the **baseline level of the outcome.** In addition, the following will be included:

Cross-sectional analyses

Cluster-level variables

Location of home (stratification variable)
Size of home (minimisation variable)
Type of home (minimisation variable)
Proportion with MMSE<20 in home at the time point at which the cross-sectional analysis is being done, as a measure of dementia in the home at that time
Baseline level of outcome for those in home and consented prior to randomisation

Individual level variables

Age at the analysis time point
Sex
Physical condition: Short Physical Performance Battery (SPPB) (at individual's baseline)
On antidepressants (Yes/No) (at individual's baseline)

Cohort analyses

Cluster-level variables

Location of home (stratification variable)
Size of home (minimisation variable)
Type of home (minimisation variable)
Proportion with MMSE<20 in home at baseline (strictly only those in the home and who provided a measure prior to randomisation as a measure of home baseline status)

Individual level variables

Baseline level of outcome
Age at home's baseline
Sex
Physical condition: Short Physical Performance Battery (SPPB)
On antidepressants (Yes/No) at baseline

Table 5: Model specifications for the different outcomes

					Covariates*											
					stratification (F)		*minimization (F)*		*allocation (F)*	*demographic (F)(I)*		*clinical (F)*	*clinical (F)*			*offset/ exposure*
Data type	Subtype	Outcome	Note	Classification#	Home location	Home ID	Home size	Home type	Int./Con.	age**	sex	baseline ***	antidepressants+ (I)	Prop. MMSE≥20 (A)	SPPB at baseline (I)	
Continuous		(b)	C	primary	✓	(R)	✓	✓	✓	✓	✓	GDS-15 (I)	✓	✓	✓	
		(c)	C	primary	✓	(R)	✓	✓	✓	✓	✓	GDS-15 (I)	✓	✓	✓	
		(r)	E	secondary	✓	(R)	✓	✓	✓	✓	✓	GDS-15 (A)	✓	✓	✓	
		(e)	E	secondary	✓	(R)	✓	✓	✓	✓	✓	MMSE (A)	✓		✓	
		(f)	C	secondary	✓	(R)	✓	✓	✓	✓	✓	MMSE (I)	✓		✓	
		(g)	E	secondary	✓	(R)	✓	✓	✓	✓	✓	SPPB (A)	✓	✓		
		(h)	C	secondary**	✓	(R)	✓	✓	✓	✓	✓	SPPB (I)	✓	✓		
		(m)	E	secondary	✓	(R)	✓	✓	✓	✓	✓	EQ-5D (A)	✓	✓	✓	
		(n)	C	secondary	✓	(R)	✓	✓	✓	✓	✓	EQ-5D (I)	✓	✓	✓	
Categorical	binary	(a)	E	primary	✓	(R)	✓	✓	✓	✓	✓	GDS-15 (A)	✓	✓	✓	
	binary	(d)	C	secondary	✓	(R)	✓	✓	✓	✓	✓	GDS-15 (I)	✓	✓	✓	
	binary	(i)	E	secondary	✓	(R)	✓	✓	✓	✓	✓	Fear of falling (A)	✓	✓	✓	
	binary	(j)	C	secondary	✓	(R)	✓	✓	✓	✓	✓	Fear of falling (I)	✓	✓	✓	
	ordinal	(k)	E	secondary	✓	(R)	✓	✓	✓	✓	✓	Pain rating (A)	✓	✓	✓	
	ordinal	(l)	C	secondary	✓	(R)	✓	✓	✓	✓	✓	Pain rating (I)	✓	✓	✓	
	ordinal	(o)	E	secondary	✓	(R)	✓	✓	✓	✓	✓	Social engage. (A)	✓	✓	✓	
	ordinal	(p)	C	secondary	✓	(R)	✓	✓	✓	✓	✓	Social engage. (I)	✓	✓	✓	
Rate		(q)	E	secondary	✓	(R)	✓	✓	✓	✓ (A)	✓ (A)		✓ (A)	✓	✓(A)	✓

* Key: (F) = fixed effect; (R) = random effect; (I) = individual level, (A) = aggregated at home level

** not measured at 6 months

*** name of baseline measure of outcome. For end-of-study outcomes, mean in home at baseline will be covariate, except proportion that feared falling.

+ on antidepressants at baseline (Yes vs. No)

** for cross-sectional analysis, use age at end of study. For cohort analysis, use age at baseline.

all change secondary outcomes will be analysed first for 6m endpoints and then for 12m endpoints.

Note: E is for end of study i.e. all residents present at the end of the study. C is for change and so is restricted to residents present at both baseline and end of study.

Outcomes:

(a) Presence of depression at 12m: whether or not resident is depressed on the basis of a GDS-15 score of five or above

(b) Change in depressive symptoms at 12 months in those residents present at baseline

(c) Change in depressive symptoms at 6 months in those residents depressed at baseline.

(d) Remission of depression at 6 months in those depressed at baseline.

(e) & (f) Cognitive function, measured using the MMSE

(g) & (h) Mobility: assessed using the Short Physical Performance Battery (SPPB) as a continuous variable.

(i) & (j) Fear of falling

(k) & (l) Pain: Participants were asked to rate their current level of pain, i.e. pain today, on a five-point numerical rating scale at 6 and 12 months

(m) & (n) Health-related quality of life (self-report & proxy)

(o) & (p) Social engagement scale

(q) Rate of peripheral fractures as a marker for injurious falls

(r) Number of depressive symptoms at 12 months (added in since version 1.0)

5.10 Multiple comparisons and multiplicity

None.

5.11 Method for handling non-adherence

There is no minimum attendance requirement at exercise classes. The reasons for absences have been documented. Homes are included in the analysis in the arm they were allocated to, regardless of adherence.

5.12 Method for handling time-varying interventions

Not applicable .

5.13 Method for handling outliers

Continuous data will be plotted prior to any modelling to assess potential data errors. These will be followed up with WCTU staff. After the model is fitted, appropriate regression diagnostics will be assessed.

5.14 Derived and computed variables

The primary outcome measure GDS-15 will define the variable 'depressed' if GDS15≥ 5.
For those who answer less than 13 items but at least 10, we apply the following rules:
depressed=Yes if GDS≥ 4 if they answered at least 11 or 12 questions.
depressed=Yes if GDS≥3 if they answered 10 items
Each item is a depressive symptom (either present (scores 1) or absent (scores 0).
If the number of items completed, n is such that: 9<n<15, the total score is to be rescaled i.e. GDS' = 15*GDS/n
MMSE (moderate-severe impairment) is defined by a total score of <20/30. If fewer than 15 items have been completed, a total score will not be computed. If between 16 and 29 items have been completed, MMSE' = 30*MMSE/n for n: 15<n<30.
Social engagement scale: if fewer than 4 items have been completed, a total score is not computed. If 4 or 5 items have been completed, SES'=6*SES/n for n = 4 or 5.
Pain now. There should only be a response to this question if the participant scored 2 or 3 on the EQ-5D pain question, otherwise set to missing. If the participant has a score of 1 on the EQ-5D pain item but 'pain now' is missing, set to 'pain now' to 0 (No).

6 DESCRIPTIVE ANALYSES

6.1 Participant flow

For the co-primary outcomes, both cluster and participant throughput will be summarised in CONSORT diagrams specific to the outcome of interest. The four main CONSORT diagrams are (a) directly assessed outcomes (cross-sectional), (b) directly assessed outcomes (cohort), (c) indirectly assessed outcomes (cross-sectional) and (d) indirectly assessed outcomes (cohort).

6.2 Representativeness of sample

N/A

6.3 Baseline comparability of randomised groups

To assess the adequacy of our randomisation process in achieving balanced groups we will compare the characteristics of RNHs (location, type and size of home) and individuals (age, sex, baseline assessment scores) in our intervention and control arms using simple descriptive statistics without recourse to hypothesis testing.

6.3.1 Demographics

Age, sex, ethnicity, age at which left school (proxy for social class) and length of residence in home will be checked descriptively.

6.3.2 Prior and concurrent medications

On antidepressants (yes/no)

6.3.3 Baseline and screening conditions

GDS-15, EQ-5D (self-report and proxy), MMSE, Barthel Index, fear of falling, current pain

6.3.4 Baseline medical history

Terminal or other serious illness are exclusion criteria

6.3.5 Baseline physical exam

Consented/assented residents were given a physical assessment (SPPB)

6.3.6 Cluster characteristics

Type (a combination of ownership and services), and size (number of beds) of RNH
Number of beds and occupancy (see table 2)

6.3.7 Characteristics of care providers where applicable

Type of RNH: local authority/voluntary/private & care/private & nursing/dementia specialist

6.4 Comparison of losses to follow-up

Proportion followed up at 6 months and at 12 months in each arm with detail presented on reasons for loss to follow-up presented in CONSORT diagrams.

6.5 Comparison of compliance to treatment and protocol

Not applicable

6.6 Emergency or accidental unblinding of randomised treatment

Not applicable

7 INTERIM ANALYSES AND SAFETY MONITORING ANALYSES

7.1 Purpose of interim analyses

To assess original sample size for adequacy.

7.2 Monitoring plan

None

7.3 Stopping rules

None

7.4 Measures taken to minimise bias

N/A

7.5 Adjustment for p-values

N/A

7.6 Interim analysis for sample size adjustment

Using data from recruitment up to November 2009, baseline data was analysed to re-estimate the power of the study. The intracluster correlation coefficient (ICC) of GDS-15 at baseline was estimated (pooled across Warwick and London). Mean cluster size was ascertained and the sample size rechecked for adequacy. See tables 3 and 4 for details.

8 ANALYSIS OF CO-PRIMARY OUTCOMES

8.1 Definition of outcome measures

i. Difference in proportion of residents depressed (defined in 2.2.1.) at 12 months between intervention and control homes.
ii. Change in number of depressive symptoms at 6 months in all those who were considered depressed at baseline, according to their GDS-15 score.
iii. Change in number of depressive symptoms at 12 months in all those who were present at baseline.

8.2 Descriptive statistics for outcome measure

By study arm, count and percentage of all participating residents depressed at 12 months, of residents present and depressed at baseline and at 12 months, and of residents present and depressed at baseline and at 6 months.

8.3 Primary analysis

To assess the difference in proportions depressed at 12 months between intervention and control homes we will use mixed effects modelling (with a random effect to account for clustering by RNH) with a binomial family and a logit link function (to handle the binary outcome) in Stata (*xtmelogit* command).

To assess the change in severity of depressive symptoms at 6 and 12 months between intervention and control homes we will use mixed effects modelling (with a random effect to account for clustering by RNH) with a Gaussian family and the identity link function in Stata (*xtmixed* command).

Each model will include the baseline measure of the outcome (except for MMSE), as well as the minimisation variables home type, home size and the stratification variable location. In addition, we will include individual baseline SPPB score and whether or not the resident was on antidepressants (cohort analyses only). We will also include, at the home level, the proportion of residents with moderate to severe cognitive impairment at baseline (i.e. with an MMSE<20).

We will examine the correlations between the covariates to assess any potential multicollinearity problem. This will be done by fitting a linear regression model and requesting the variance inflation factors (VIFs). A VIF>10 indicates severe multicollinearity. VIFs between 5 and 10 indicate some problem.

For the outcomes which indicate a change in severity of depression for those depressed at baseline, we will use a similar strategy, but measurements at baseline and outcome will be on the same individual. We will therefore introduce baseline levels of depression severity as a covariate at the individual level.

All covariates listed in section 5.9 will be entered into the model at the same time and none will be removed unless there is evidence of multicollinearity, in which case, the collinear covariate providing the best fit will be retained.

The regression coefficient for continuous outcomes, odds ratio for binary outcomes, proportional odds ratio for ordered outcomes and incidence rate ratio for rate outcomes, of the treatment (intervention vs. control) variable, and its 95% confidence interval, will be reported along with the p-value.

8.4 Assumption checks

Appropriate model diagnostics will be checked: e.g. normality (continuous outcomes only) and homoscedasticity of residuals and normality of random effects.

8.5 Other analysis supporting the primary (inc. sensitivity analyses)

Residents joining post-randomisation

There are two groups of residents to consider here. By far the largest group is new residents who join the study at 3, 6 or 9 months post-randomisation. The second group comprises a small number of residents who were present in the home at baseline and for whom some data were collected soon after their home was randomised. For example, proxy EQ-5D just before randomisation and a self-completed one just after randomisation.

In the cross-sectional analysis inclusion criteria will be based on length of time resident in home; thus individuals from these two groups may be included in the main analysis. A sensitivity analysis will exclude these individuals.

In the cohort analyses these two groups will not be included in the main analysis. They will be included in a sensitivity analysis providing that they have been resident in the home prior to pre-determined cut-offs set at about 3 months prior to each outcome end point. Their baseline will be the earliest time point at which they are assessed. The cohort analysis has endpoints at 6 and 12 months. This will provide an analysis population that is comparable to that in the health economics analysis.

Missing values
For cohort study outcomes:
Assuming that the linear predictor is correct, the mixed models will provide efficient and unbiased effect estimates despite missing values in covariates, assuming that the missingness mechanism is missing at random (MAR).

After the analyses set out in this plan have been carried out, KDO will perform analyses under various missing data assumptions, to be detailed in a separate analysis plan.

9 ANALYSIS OF SECONDARY OUTCOMES

9.1 Definition of outcome measures

- The proportion of depressed residents, with 'depression' measured prior to randomisation, that experience remission from their 'depression' six months after their home was randomised.
- Cognitive function: Mini Mental State Examination as a continuous variable
- EQ-5D as a continuous variable: proxy and self-reported
- Mobility: Short Physical Performance Battery (SPPB) as a continuous variable.
- Falls:

  (a) Binary fear of falling: "Are you afraid of falling" yes/no response from participants.
  (b) Rate of peripheral fractures as a marker for injurious falls. We will identify these from the RNH (home aggregate level) and hospital records (individual level).
- Pain: ordinal 5-point rating scale at each follow-up.
- The Social Engagement Scale uses six items from the Minimum Data Set residential assessment instrument (MDS-RAI).
- Medication use: For antidepressants, other psychoactive drugs, analgesics and NSAIDs we will convert these into the number of defined daily doses (DDD) used over one year (http://www.whocc.no).
- Hospital admissions: cause and duration of any hospital admissions during the study period from participants' hospital records. We will code these admissions into Diagnosis Related Groups to identify any peripheral fractures.

9.2 Descriptive statistics for outcome measure

Continuous outcome: n, mean, SD, median, 10^{th} and 90^{th} centiles, by arm.
Dichotomous outcome: n (%), by arm.

9.3 Secondary analysis

The analysis strategy is the same as for the co-primary outcomes.

- Remission of depression: reduction in proportion of participants depressed in the intervention arm. A mixed effects logistic model will be defined. This will be fitted using the *xtmelogit* command.

- Cognitive function Mini Mental State: We will compare the rate of change in MMSE cluster mean at baseline and at end of intervention in exercise and control homes. We could construct linear mixed effects models with MMSE at 0, 6, and 12 months, with random effects for RNH. This will be fitted using the *xtmixed* command.

- EQ-5D: we will use linear mixed models, with RNH as a random effect. This will be fitted using the *xtmixed* command.

- Mobility: (SPPB) we will use linear mixed models, with RNH as a random effect. This will be fitted using the *xtmixed* command.

- Falls
  Binary fear of falling: reduction in proportion of participants declaring fear of falling in the intervention arm. A mixed effects logistic model will be defined. This will be fitted using the *xtmelogit* command.

- Pain: 5-point scale. Ordinal mixed effects logistic regression (either in Stata using the *gllamm* command or in MLWiN.)

- The Social Engagement Scale uses six items from the Minimum Data Set residential assessment instrument (MDS-RAI). Ordinal mixed effects logistic regression (either in Stata using the *gllamm* command or in MLWiN.)

  The following outcomes will not be modelled. However, medication use and hospital admissions are important for the health economic analysis (see separate analysis plan).

- Medication use (descriptive statistics only)
- Hospital admissions (descriptive statistics only)
- All-cause mortality.

9.4 Assumption checks

Appropriate model diagnostics will be checked: normality (continuous outcomes) and homoscedasticity of residuals, normality of random effects.

9.5 Other analysis supporting the secondary (inc. sensitivity analyses)

See section 8.5 for details.

10 SAFETY AND TOLERABILITY ANALYSES

10.1 Drug exposure

N/A

10.2 Adverse events

Data on adverse events was not collected.

10.3 All adverse events

N/A

10.4 Adverse events leading to withdrawal

N/A

10.5 Serious adverse events

According to the 'adverse event reporting protocol', an SAE is any untoward or medical occurrence that: results in death or is life threatening *or* requires hospitalisation *or* results in persistent disability of incapacity *or* requires medical intervention to prevent one of these (or is otherwise considered medically significant by the investigator).

SAEs can be classified as directly attributable (i.e. occur during a study assessment or getting to or from or occur during a group-activity session), or indirectly attributable (i.e. at any other time between randomisation and end of follow-up).

Table 6 Classification of anonymised fracture-related deaths, and fractures data.

Deaths* and Fractures (indivdiual level data)	Attribution of SAE	
	direct	indirect
study participant (intervention)	A, E	otherwise
non study participant (partaking in exercise)	E	otherwise
study particpant (control)	A	otherwise

A = during study assessment
E = during exercise
otherwise = not (A or E)
* fracture-related deaths (ICD10 chapter S)

Difference in rates of peripheral fractures between intervention and control homes (mixed effects Poisson model with resident-time in study as the exposure variable). This will be fitted using the *xtmepoisson* command.

Reporting of cause of death data

All deaths in residents will be reported and documented. We will have data on deaths amongst all residents from all homes. This will not provide accurate dates of death and characteristics of residents. We will conduct a simple comparison of total death rate in the two trial arms. For those who have consented/assented to have their data used, we will have exact date of death from death certification data, plus age and gender and all the other baseline variables. For this group we will construct Kaplan-Meier plots and carry out log-rank tests to see if there is any difference in all cause mortality over time. We will also consider using Cox's proportional hazards models (*streg* command) or logistic regression to allow for covariates depending on the number of deaths.

All cause of death data will be based on ICD10 coding provided by MRIS. These deaths data are provided as the underlying cause of death and also as the factors contributing to death.

This is a safety analysis that will be presented for all subjects contributing data to the study.

We will present the number of deaths for which any injury, i.e. any underlying of contributory cause of death includes an ICD10 code that starts with an 'S', has contributed to the cause of death. We will present these as fracture-related deaths per year of exposure. We will estimate difference in fracture related death rate within a 95% confidence interval.

For the primary cause of death outcome the underlying cause of death from the MRIS data will be used. This will presented as number of deaths during study period, subdivided by ICD chapters; i.e. the first character in the ICD code, between each group. No statistical analyses will be presented.

10.6 Clinical laboratory evaluations

N/A

11 SUBGROUP ANALYSES

11.1 Definition of outcome measure

N/A

11.2 Definition of subgroups

N/A

11.3 Sample size justification for the subgroup analysis

N/A

11.4 Descriptive analysis for subgroups

N/A

11.5 Method of analysis

N/A

12 AMENDMENTS TO VERSION 1.0

Barthel Index replaced by SPPB as a measure of physical function because Barthel Index was missing for all residents in one home at baseline.
If one or two items are missing in the Social Engagement Scale, the score will be pro-rated according to the rule written in 5.14.
To avoid having a category with only one member, local authority homes and voluntary homes will be collapsed into one category for the purpose fitting the 'home type' design variable in the analysis.

13 REFERENCES

Rothera IC, Jones R, Harwood R, Avery AJ, Waite J. Survival in a cohort of social services placements in nursing and residential homes: factors associated with life expectancy and mortality. Public Health 2002;116:160-5.

Underwood M, Eldridge S, Lamb S, et al. The OPERA trial: protocol for a randomised trial of an exercise intervention for older people in residential and nursing accommodation *Trials* 2011, 12:27 doi:10.1186/1745-6215-12-27

14 APPENDIX

Where scoring of questionnaires is required for the analysis they should be attached here