Bristol Randomised Trials Collaboration (BRTC)

3D study

Statistical Analysis Plan

Version 1.0 (May 19th 2017)

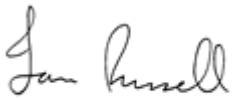
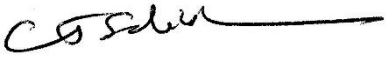
The following people have reviewed the Statistical Analysis Plan and are in agreement with the contents			
Name	**Role**	**Signature**	**Date**
Bryar Kadir Sara Brookes	Authors	Bryar Kadir Sara Brookes	19/05/17
Ian Russell (DMC)	Chair DMC		
Chris Salisbury	Chief Investigator		22/05/17

Table of Contents

List of tables

Abbreviations

3D	The name of the intervention under evaluation; '3D' stands for **D**imensions of health, **D**epression and **D**rugs
AE	Adverse Events
BRTC	Bristol Randomised Trials Collaboration
CACE	Complier Causal Average Effect
CARE	Consultation and Relational Empathy questionnaire
CONSORT	Consolidated Standards of Reporting Trials
COPD	Chronic Obstructive Pulmonary Disease
CI	Confidence Interval
CKD	Chronic Kidney Disease
CRCT	Cluster Randomised Controlled Trial
CVD	Cardiovascular Disease
DMC	Date Monitoring Committee
EQ-5D	Euroquol 5D Questionnaire
GP	General Practitioner
HADS	Hospital Anxiety Depression Scale
ICC	Intra Cluster Coefficient
IQR	Inter-Quartile Range
LTC	Long Term Condition
LTC6	Six item Long-Term Conditions questionnaire
MICE	Multiple Imputation Chained Equations
MID	Minimally Important Difference
MMAS	Morisky Medication Adherence Scale
PACIC	Patient Assessment of Chronic Illness Care measure
QOF	Quality and Outcomes Framework
SAE	Serious Adverse Event
T0	Baseline
T1	9 month post randomisation
T2	15 month post randomisation
TMG	Trial Management Group
TSC	Trial Steering Committee
UKCRC	UK Clinical Research Collaboration
WISE	Whole system Informing Self-management Engagement intervention

1. INTRODUCTION & PURPOSE

This document details the rules proposed and the presentation that will be followed, as closely as possible, when analysing and reporting the main results from the 3D study: A pragmatic cluster randomised controlled trial of a management intervention for patients with multi-morbidity in general practice.

The purpose of the plan is to:

1. Ensure that the analysis is appropriate for the aims of the trial, reflects good statistical practice, and that interpretation of *a priori* and *post hoc* analyses respectively is appropriate.
2. Explain in detail how the data will be handled and analysed to enable others to perform the actual analysis in the event of sickness or other absence.
3. Protect the project by helping it keep to timelines and within scope.

Additional exploratory or auxiliary analyses of data not specified in the protocol are permitted, but fall outside the scope of this analysis plan (although such analyses would be expected to follow Good Statistical Practice).

The analysis strategy will be made available if required by journal editors or referees when the main papers are submitted for publication. Additional analyses suggested by reviewers or editors will, if considered appropriate, be performed in accordance with the Analysis Plan, but if reported the source of such a *post-hoc* analysis will be declared.

Editorial changes

Amendments to the statistical analysis plan will be described and justified in the final report of the trial and in Section 9 of this document.

2. STUDY DESIGN AND OBJECTIVES

The information in these sections has been extracted from the protocol (version 7) and the published trial protocol[1] and is presented here to place the analysis plan within the context of the trial aims and methods. More detail is provided in the protocol version 7 and the published protocol[1].

2.1. Trial aims and objectives

2.1.1. Primary aim

To optimise, implement and evaluate an intervention to improve the management of patients with multi-morbidity in general practice.

2.1.2. Objectives

1. To optimise the intervention through piloting in three practices
2. To assess, through a **cluster randomised controlled trial (CRCT)**, the impact of the intervention on health-related quality of life, illness burden, treatment burden, and patient experience as well as carer's burden and quality of life.
3. To assess the cost effectiveness of the intervention with an **economic evaluation**.
4. To explore, through a mixed methods process evaluation, how and to what extent the intervention was implemented, and how/why the intervention was/not beneficial. To explore the advantages /disadvantages of different models of care for patients with comorbidity and to characterise usual care and explore any changes to management practices over the duration of this study in usual GP practice.

This analysis plan relates to objectives 2 and 3 only. Methods and analyses relating to objective 4 are detailed in the process evaluation protocol paper[2].

2.2. Trial design and configuration

This is a multi-centre, pragmatic, two-arm, practice-level CRTC comparing a new approach to the management of multi-morbidity in general practice versus usual care, with a parallel economic analysis of cost effectiveness and a mixed methods process evaluation.

2.3. Trial centres

Recruiting centres are based in Bristol, Manchester and Glasgow. GP practices will be recruited from in and around Bristol, Manchester and Ayrshire and Arran.

2.4. Eligibility criteria

2.4.1. Inclusion criteria for practices

General practices with the following criteria will be approached:

- Minimum of three GP partners
- Minimum practice list size of 4,500 patients
- Uses EMIS GP computer system

2.4.2. Exclusion criteria for practices

There are no exclusion criteria at cluster level.

2.4.3. Inclusion criteria for patients

- Aged 18 or over (on date of invitation to participate)
- Three or more Long Term Conditions (LTC) from the following list: cardiovascular disease (CVD) or chronic kidney disease (CKD) (including coronary heart disease, hypertension, heart failure, peripheral arterial disease); stroke; diabetes; COPD or asthma; epilepsy; atrial fibrillation; severe mental health problems; depression; dementia; learning disability and rheumatoid arthritis.

2.4.4. Exclusion criteria for patients

- Life expectancy less than 12 months
- Serious suicidal risk
- Known to be leaving practice within 12 months
- Cannot complete a questionnaire in English (alone or with help)
- If actively taking part in other research involving extra visits to GP or other health services
- Adults lacking capacity to consent (Scotland only)

2.5. Description of interventions

The 3D intervention is a new approach to the management of patients with multi-morbidity. It is a complex intervention with four main components:

- **Identification and prioritisation of patients with multi-morbidity** – identified patients will be given a '3D' card and their GP records 'flagged'.
- **Improving patient-centred care** – each '3D' patient will be allocated a named usual GP and usual practice nurse who will have responsibility for co-ordinating their care. They will be offered longer appointments with their usual doctor or nurse when possible.
- **Reducing the burden of illness and treatment** – '3D' patients will be offered a comprehensive assessment every 6 months instead of separate reviews for each of their LTCs. Each 3D assessment consists of 2 appointments approximately 1 week apart. At the first appointment, the patients' usual nurse will complete a bespoke computerised template to address all of the 3D elements (Dimensions of health, Depression and Drugs), collect relevant data in relation to the patient's combination of LTCs, and organise necessary tests. At the second appointment, the usual GP will review all the information, conduct a thorough review of medication and agree a written care plan with the patient for them to take away. Before the GP completes the 3D assessment a pharmacist will review each 3D patient's medication list and provide recommendations for the GP. This pharmacist review will be undertaken once during the year for each patient.

- **Improving integration** – each practice has a designated general physician who is readily available to discuss multi-morbidity patients with complex needs and help co-ordinate hospital investigations.

To ensure the intervention is effectively implemented, it will be incentivised as if it were an Enhanced Service or included in the Quality and Outcomes Framework (QOF), with payment against targets for completion of two 3D assessments per annum.

Practices allocated to the control arm will continue care as usual. In most practices this will mean patients are recalled to different clinics to see different practice nurses to review each of their long-term conditions.

2.6. Randomisation procedures

To minimise post-randomisation selection bias, practices will not be randomised until after patients have been identified and after the initial patient invitations have been mailed.

Practices will be randomised using an algorithm written in advance by the Bristol Randomised Trials Collaboration (BRTC, UKCRC registration ID: 2) on a 1:1 ratio to receive either the intervention or continue care as usual (control group). Randomisation will be stratified by recruiting centre (Bristol, Manchester, and Glasgow) and minimised by practice deprivation level and practice size. Practices within each area will be randomized using a block size of two (one randomized to the 3D intervention and the other to usual care), to ensure balance across the treatment arms given the relatively small number of practices. Within each centre, each block of two practices will be randomized at the same time in the following way.

Within each centre, the initial block of two will be randomized using simple randomization, such that one is allocated to intervention and the other control. For each subsequent block of practices, an algorithm (written within Stata specifically for this study) will determine the allocation of the two practices which creates the best balance in terms of size and deprivation and then weights the randomisation in favour of this allocation (rather than being deterministic); the weights being determined by the degree of imbalance in terms of size and deprivation (see Table 1 and example below).

Example use of Table 1: Suppose the first practice in the next block of two is allocated to control and the second to 3D (denoted allocation 01 in Table 1) and that this would lead to an absolute difference in median practice size between the two treatment groups of 327, whereas if the first practice is allocated to 3D and the second to control (allocation 10), the absolute difference in median practice size is 116. Then the difference in imbalance (allocation 01 minus allocation 10) in terms of practice size is +211, a greater imbalance when the allocation is 01. Suppose also that allocation 01 would lead to an absolute difference in median deprivation score between the two treatment groups of 3 whereas allocation 10 would lead to an imbalance of 9. Then the difference in potential imbalance (allocation 01 minus allocation 10) would be -6, a greater imbalance when the allocation is 10. From Table 1, considering potential imbalance in both size and deprivation, this would result in a weighting of 0.65 in favour of allocation 01.

Table 1: Randomisation weightings (in favour of allocation 01[a]) for each block of two practices

Practice size Difference in imbalance (allocation 01 minus allocation 10)[a]	**Deprivation score** Difference in imbalance (allocation 01 minus allocation 10)[a]						
	≤-12	-11 to -8	-7 to -4	-3 to 3	4 to 7	8 to 11	≥12
≤-900	0.80	0.80	0.80	0.8	0.75	0.65	0.50
-899 to -600	0.80	0.80	0.80	0.75	0.65	0.50	0.35
-599 to -300	0.80	0.80	0.75	0.65	0.50	0.35	0.25
-299 to 299	0.80	0.75	0.65	0.50	0.35	0.25	0.20
300 to 599	0.75	0.65	0.50	0.35	0.25	0.20	0.20
600 to 899	0.65	0.50	0.35	0.25	0.20	0.20	0.20
≥900	0.50	0.35	0.25	0.20	0.20	0.20	0.20

[a]Allocation 01 - first practice in the block of two is allocated to control and second practice allocated to 3D; allocation 10 – first practice allocated to 3D and second practice to control

	Randomisation of two practices weighted in favour of allocation 01
	Randomisation of two practices weighted against allocation 01 (hence, in favour of allocation 10)

2.6.1 Allocation concealment

Practices within each of the three areas will be randomized using a block size of two. The trial manager (not involved in practice recruitment) will provide the senior statistician with the two practice IDs, size and deprivation scores. This information will then be given to the statistician only after practice level consent and after eligible patients have been identified and invited to take part. Allocations will be generated via a pre-written Stata do file. Whilst the study team are aware of the consistent block size of two, randomizing the two practices together will enable those recruiting practices to remain unaware of the next allocation. The statistician will inform the trial manager of the allocation of the two practices via email; the individual recruiting practices will then inform the practices.

2.7. Sample size and justification

The study is designed to detect an effect size of 0.274 standard deviations in the primary outcome of the EQ5D-5L. Data about the variability of the new 5 level (5L) version of the EQ5D is currently more limited than for the well-established 3 level (3L) version. The standard deviation of the EQ5D-3L in the UK general population is 0.23, rising to 0.27 in the oldest respondents (aged over 75)[3]. Hence an effect size of 0.274 would equate to a detectable difference of (0.274*0.27) = 0.074 on the EQ5D-3L, previously deemed to be the minimum important difference (MID)[4]. Although less data is available about the variability in

the 5L version of the EQ5D than the 3L version, it seems wise to use this latest version of the EQ5D as it is likely to have greater sensitivity to change.

Based on data available from our previous studies, we estimate that 2.3% of adult patients will have multi-morbidity in terms of three or more LTCs as defined in this study. This equates to about 108 patients in an average sized practice of 6000 patients i.e. 3456 potentially eligible patients in 32 practices. Assuming 40% of patients agree to participate (n=1382), 80% are followed up to 12 months, and an Intra-Cluster Coefficient (ICC) of 0.03 for clustering at the practice level (based on the WISE trial)[5] this sample provides around 90% power at a 5% significance level to detect an effect size of 0.274 standard deviations in the EQ5D measure between the intervention and control groups.

Since the start of this study, we are aware of two studies that have been published determining an MID for the EQ-5D-5L based on UK data. Nolan et al.[6] published EQ5D-5L data from 616 COPD outpatients (mean age 70.4 years), reporting a standard deviation of 0.24 (consistent with the EQ5D-3L and the above sample size calculation). They used distribution- and anchor-based methods to determine a MID for COPD of 0.051 (95% CI 0.037 to 0.063). McClure et al.[7] used a simulated-approach based on instrument-defined health transitions and identified an MID for England of 0.063 (SD 0.013). The sample size calculation for the 3D study was determined to detect a difference of 0.074 (based on the EQ5D-3L - the best estimate for the EQ5D-5L at the time). Interpretation of the 3D trial findings will also include consideration of alternative MIDs (such as Nolan and McClure) suggested in the literature since the start of the trial, with the acknowledgement that the study may be underpowered to detect an MID smaller than 0.074.

2.8. Blinding and breaking of blind

It is not possible to mask participants or health care professionals to the group allocation of their practice. It is not possible to keep all members of the study team blind, however, efforts will be made to blind members that can be blinded, this will include the junior trial statistician, who will carry out the analysis of outcomes.

2.9. Trial committees

The Trial Management Group (TMG) will meet regularly (every 6-8 weeks) to ensure the three study centres are working consistently, meeting study targets and adhering to the study protocol. The group will consist of the CI, Trial Manager, PI and researchers from each of the recruiting centres with input from other members of the research team where necessary. Regular progress reports regarding study recruitment, retention, issues or complaints and adverse events will be reported and discussed.

An independent Trial Steering Committee (TSC) will be convened comprising of an external academic chair (who is also an academic GP), at least two other independent members (which will include an independent statistician and an independent clinician with relevant experience or interests), two patient representatives, the CI, PIs and other key members of the study research team. The TSC will meet at least annually (face-to-face or by teleconference) or more frequently at the request of the chair. The TSC will provide external supervision to the study and monitor the overall trial progress, adherence to the protocol and the implications of any new information (e.g. research articles or policy changes).

A Data Monitoring Committee (DMC) will comprise of an independent chair and at least two other independent members including an independent statistician and a clinician with relevant interests. The CI, Trial Manager and Trial statistician will report to the DMC. The remit of the DMC is to monitor the trial data, in particular to quality control and quality assurance of data collected; the progress of the trial, including recruitment and retention rates and adherence to the trial protocol. A key role is to ensure that the dignity, rights, safety and well-being of the study participants are maintained at all stages of the trial. All adverse events will be reported to the committee which can have direct access to source data and documentation. Where possible the DMC will convene prior to the TSC and will report their recommendations to the TSC.

2.10. Outcome measures

2.10.1. Primary outcome

EQ5D-5L descriptive system (measured at 15 months): The descriptive system comprises 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each represented in a single question. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box against the most appropriate statement in each of the 5 dimensions. This decision results in a 1-digit number expressing the level selected for that dimension. The digits for 5 dimensions can be combined in a 5-digit number. These are then transformed onto a single continuous scale, using a value set, summarising the respondent's health state. The scale is anchored at 1 (full health) and 0 (a state equivalent to dead); a health state considered to be worse than death is given values less than 0.

2.10.2. Secondary outcomes

Whilst the primary time point for analyses is 15 months the 3D trial will also consider the potential effectiveness of the 3D intervention at 9 months. Secondary analyses will consider the primary and secondary outcomes at 9 months, rather than employing repeated measures analyses which would estimate the effectiveness of the 3D intervention over the duration of the study and examine if there was a differential effect across time. It is plausible that the 3D intervention will be effective in the short-term only with any effects disappearing by 15 months, this may still warrant its implementation into primary care practice and so its effectiveness at 9 months should be formally tested within this study.

The table below lists the secondary outcomes to be considered in analyses and the time points at which they will be measured (T0=baseline; T1=9 months; T2=15 months). Analyses for each secondary outcome will include adjustment for the relevant measure at baseline (T0).

Table 2: Secondary outcomes to be collected in the 3D study, to be compared between trial arms

Secondary outcome	Source	Scale	Time point		
			T0	T1	T2
Experience of holistic patient-centred care					
CARE measure of relational empathy (GP)	Questionnaire - 10 items	10-50	✓	✓	✓
CARE measure of relational empathy (nurse)	Questionnaire - 10 items	10-50	✓	✗	✓
Care related to patients' priorities (LTC6)	Questionnaire – 1 item	1-4	✓	✓	✓
Care which is joined up (LTC6)	Questionnaire – 1 item	1-4	✓	✓	✓
PACIC measure of chronic disease management	Questionnaire - 20 items	1-5	✓	✓	✓
Overall satisfaction	Questionnaire – 1 item	1-5	✓	✓	✓
Burden of illness measures					
EQ5D-5L	Questionnaire -5 items		✓	✓	✗ [1]
Self-rated health	Questionnaire – 1 item	1-5	✓	✓	✓
Bayliss measure of illness burden	Questionnaire – 27+ items	0-145[2]	✓	✓	✓
HADS Anxiety score	Questionnaire - 7 items	0-21	✓	✓	✓
HADS Depression score	Questionnaire - 7 items	0-21	✓	✓	✓
Burden of treatment					
Multimorbidity Treatment Burden Questionaire[3]	Questionnaire - 10 items	0-100	✓	✓	✓
Medication adherence: MMAS-8	Questionnaire - 8 items	0-8	✓	✓	✓
Number of drugs prescribed[4]	Practice records	≥0	✓	✗	✓
Number of high risk prescribing indicators	Practice records	≥0	✓	✗	✓
Process measures[5]					
Continuity of care - COC index	Practice records	0-1	✓	✗	✓
Continuity of care - Visit Entropy	Practice records	0–log2(1/k[6])1	✓	✗	✓
Quality of disease management[7]	Practice records	0-100	✓	✗	✓
Number of hospital admissions	CSU data[8]	≥0	✓	✗	✓
Number of outpatient attendances	CSU data[8]	≥0	✓	✗	✓

[1] EQ5D-5L is also measured at 15 months; this is the primary outcome for the study not a secondary outcome

[2] No ceiling score, thus no maximum, respondent can add extra items (note: in 3D questionnaire only 2 extra conditions can be added hence maximum is inferred).

[3] Questionnaire developed for this trial

[4] Number of <u>different types</u> of drugs prescribed in the previous 3 months (different prescriptions of same drug or different dosages/formulations of same drug will <u>not</u> be counted as additional prescriptions).

[5] Whilst process measures, these outcomes are included as secondary outcomes as a primary aim of the study was to consider these as measures of whether the 3D intervention is effective.

[6] k is the total number of possible providers

[7] Percentage of relevant Quality and Outcomes Framework (QOF) indicators met by each patient.

[8] CSU data if available, otherwise patient questionnaire

Table 3: Carer secondary outcomes to be collected for 3D study, to be compared between trial arms

Secondary outcome	Source	Scale	Time point		
			T0	T1	T2
Carer experience scale	Questionnaire - 6 items	0-100	✓	✓	✓
EQ5D-5L carer score[1]	Questionnaire - 5 items	-1 to 1	✓	✓	✓
Multimorbidity Treatment Burden Questionnaire[1,2]	Questionnaire – 16 items[3]	Scale still under development	✓	✓	✓

[1] Not listed in the trial registry

[2] Questionnaire developed for this trial

[3] The scoring system for the Multimorbidity treatment Burden Questionnaire is still under development; it is possible that not all 16 items will be used in the final score.

The number of carers with completed questionnaires is small, hence analyses will be exploratory in nature.

2.10.3 Process measures

Table 4 presents the process measures to be considered in analyses and the time points at which they will be measured.

Table 4: Patient level process of care measures to be collected in the 3D study, to be compared between trial arms

Process measure	Source	Scale	Time point		
			T0	T1	T2
Number of primary care consultations with GP	Practice records electronic extract	≥0	✓	✗	✓
Number of primary care consultations with nurse	Practice records electronic extract	≥0	✓	✗	✓
Mean duration of face to face consultations in surgery with GP[1]	Practice records electronic extract	≥0 (mins)	✓	✗	✓
Mean duration of face to face consultations in surgery with nurse[2]	Practice records electronic extract	≥0 (mins)	✓	✗	✓
Number of different review consultations[3]	Practice records electronic extract	≥0	✓	✗	✓
At least one review received	Practice records electronic extract	0 (No), 1 (Yes)	✓	✗	✓

[1] Face to face consultations only because duration of telephone consultations and home visits are not reliably recorded

[3] Include all the chronic disease review codes including codes for 3D nurse review and GP review (nurse and GP reviews counted as separate reviews). Within each patient sort by date, and delete duplicate codes on one date, so that if they have several different diseases coded on the same day this just counts as one review. Then count how many review consultations each patient had.

The number of chronic disease reviews for diabetes (based on diabetic foot risk assessment), asthma, COPD, dementia, mental health and rheumatoid arthritis will be summarised for each treatment group. Percentages will be presented using the number of patients having that disease as the denominator. No hypotheses tests will be carried out comparing the groups.

Table 5 presents additional process measures to be reported for the 3D intervention arm only. There will be no comparative analyses.

Table 5: Patient level process of care measures to describe implementation of the intervention (descriptive data, collected in the intervention arm only)

Process measure	Source	Scale	Time point		
			T0	T1	T2
Number of nurse 3D reviews	Practice records	0, 1, 2	✗	✗	✓
Number of GP reviews	Practice records	0, 1, 2	✗	✗	✓
Compliance[1]	Practice records	None, partial, full reviews	✗	✗	✓
Most important problem noted	Practice records	0 (No), 1 (Yes)	✗	✗	✓
EQ5D pain question noted	Practice records	0 (No), 1 (Yes)	✗	✗	✓
PHQ9 entered	Practice records	0 (No), 1 (Yes)	✗	✗	✓
Medication reviewed by Pharmacist (at least one comment entered)	Practice records	0 (No), 1 (Yes)	✗	✗	✓
Medication adherence noted	Practice records	0 (No), 1 (Yes)	✗	✗	✓
First patient goal noted	Practice records	0 (No), 1 (Yes)	✗	✗	✓
First plan noted ('what patient can do')	Practice records	0 (No), 1 (Yes)	✗	✗	✓
First plan noted ('what GP can do')	Practice records	0 (No), 1 (Yes)	✗	✗	✓
Patient agenda printed[2]	Practice records	0 (No), 1 (Yes)	✗	✗	✓
3D plan printed[2]	Practice records	0 (No), 1 (Yes)	✗	✗	✓
Number of times hospital physician was contacted	Physician records	≥0	✗	✗	✓

[1] Compliance defined as: full – two GP 3D appointments and two nurse 3D appointments; 'partial' – at least one GP or nurse 3D appointment; and 'none' – no GP 3D appointment and no nurse 3D appointment (see section 6.5.3).

[2] not available in practices in Scotland

3. GENERAL ANALYSIS CONSIDERATIONS

3.1. Analysis populations

Full analysis set: All patients that consent to take part in the study within a randomized practice. These participants (practices) will be analysed in the groups to which they were allocated, disregarding protocol deviations or non-compliance. Missing data will be imputed using multiple imputation modelling (see section 3.3). Patients who have not completed the EQ5D-5L because they are deceased will be recorded as having a value of 0.

Complete cases set: All patients that consent to take part in the study within a randomized practice. These participants (practices) will be analysed in the groups to which they were allocated, disregarding protocol deviations or non-compliance. Missing data will not be imputed. Patients who have not completed the EQ5D-5L because they are deceased will be recorded as having a value of 0.

3.2. Derived variables

3.2.1 Primary outcome

A UK value set for the EQ5D-5L is now available[8], along with a Stata do-file to transform the five EQ5D responses into a single value summarising the respondents' health state.

3.2.2 Secondary outcomes

CARE – 10 item questionnaire (each scored between 1-5), total score is summation of individual scores.[9]

PACIC – 20 item questionnaire (each item scored 1-5), overall score is an average of all 20 items.[10]

EQ5D-5L (at 9 months): As calculated for patients in primary outcome.

Bayliss – For each of 27 chronic conditions respondents select those that they experience and rate each selected condition on a five-point scale from 1 (interferes with daily activities "not at all") to 5 (interferes with daily activities "a lot"). Respondents are additionally allowed to add medical conditions not already on the list. The overall score representing level of morbidity is then the sum of conditions selected weighted by the level of interference assigned to each (that is, the sum of the interference scores).[11]

HADS: Anxiety score – simple addition of the relevant 7 questions; Depression score – simple addition of the relevant 7 questions.[12]

Multimorbidity Treatment Burden Questionnaire – 10 items each scored 0-4. Total score is calculated by calculating the average score for each patient and then multiplying by 2.5 to get a value 0-100.

MMAS – 8 item questionnaire, with 7 questions being binary outcome assigned values 0 or 1, and a single question has values 4, 3, 2, 1 or 0. Questions 1-4 and 6-7 are coded 0 for 'yes' and 1 for 'no'; question 5 is coded 0 for 'no' and 1 for 'yes'. To get total score, add 7 binary questions together, and divide value of other single question by 4, and add them together to get a score between 0-8.[13]

Number of high risk prescribing indicators – This is the number of adverse warnings triggered and can be from 0 upwards.[14]

Continuity of care: Visit entropy will be used to measure the continuity of care.

A patient's visit to a healthcare provider can be considered a discrete random variable X. Then X can take on k distinct levels, one for each healthcare provider the patient could visit. It has a probability distribution function p(x) that represents the probability of visiting each healthcare provider.

Visit Entropy H(X) of a discrete random variable X can be calculated as[15]:

$$\mathrm{H}(\mathbf{X}) = -\sum_{i=1}^{k} p(x_i) log_2\, p(x_i)$$

and the probability of visiting the i^{th} provider is estimated as:

$$\hat{p}(x_i) \approx \frac{n_i + {}^{1}/_{k}}{N+1}$$

Where n_i is the number of observed visits to the i^{th} provider, k is the total number of possible providers, and N is the total number of observed visits.

H(X) approaches its minimum value of zero when a patient has perfect continuity of care, visiting only their primary physician, and approaches its maximum when there is no continuity of care.

Visit entropy is a relatively new measure, and is less well recognised than other measures for continuity of care. For this reason, continuity of care index (COCI) will also be considered for comparison.

Continuity of care index (COCI) formula[16]:

$$COCI = \frac{(\sum_{j=1}^{M} n_j^2) - N}{N(N-1)}$$

Where N is the total number of visits; n_j is the number of visits to the jth different provider, where j= 1, 2, 3, … M, and M is the number of potential available providers.

Carers' experience scale – 6 item questionnaire (each item has 3 possible responses). Preference-based index values are available to transform the 6 responses to a profile measure value between 0 and 100[17].

EQ5D-5L (carers): As calculated for patients in primary outcome.

Multimorbidity Treatment Burden Questionnaire (carers) – development of final scoring system still under-development.

Quality of disease control – This is based on the Quality and Outcomes Framework (QOF) indicators and uses the 'patient average' method of Reeves *et al*[18]. It will be measured as a percentage for each individual patient, where it represents the percentage of QOF chronic disease management indicators that apply to that patient which were successfully met.

3.3. Procedures for missing data

Missing data may arise as some participants may not return their questionnaires. It is anticipated that proportions with missing data will be similar between the two randomization arms but this will be examined and reported. Baseline characteristics will be compared between participants with and without 15-month follow-up data. In all tables missing data will be indicated by footnotes.

The primary analysis for EQ5D-5L (section 6.2) will include the full analysis set (section 3.1), including all patients in the groups to which they were allocated and imputing missing data. Reasons for missingness will be explored. Missing data will be imputed using multiple imputation techniques such as multiple imputation by chained equations (MICE). Imputation models will include baseline, 9 month and 15 month EQ5D-5L data (as available), intervention arm, stratifying/minimisation variables, as well as other variables such as baseline covariates and auxiliary covariates that are informative of missingness. To allow for clustering in Stata, imputations will be performed separately for each practice if possible. The same imputation models will be used for the primary analysis and the economic evaluation if possible. The influence of missing data on the primary analysis will be investigated in sensitivity analyses using complete case data only (section 6.5).

The analyses of secondary outcomes will use complete case data only. The only exception to this is for secondary outcomes derived from several items within a validated questionnaire, where questionnaire guidelines specifically state that an overall score can be derived even in the presence of missing data for one (or more) item(s). For example, for HADS anxiety and depression scores, questionnaire guidelines state that the score for a single missing item from a subscale is inferred by using the mean of the remaining six items; however, if more than one item is missing from that subscale the overall score is missing (and the observation excluded from the analysis). The numbers (percentage) of missing data will be presented in tables of secondary outcomes.

3.4. Study centre effects

Randomisation is at the general practice (centre) level; the effect of practice will be taken into account as a random effect in multi-level regression models.

3.5. Outliers

Data will be checked for validity, each variable will be examined separately, and any outliers (> 3SD of the mean) will be checked for entry errors. Where no error is found, the variable will be checked for concordance with other variables, differences will be noted. We will also examine for influential observations (Cooks distance[19]) in the main analysis models. Outliers and influential observations will be noted. Sensitivity analyses removing outliers will be conducted.

3.6. Data cut-off

The cut-off for outcome data to be included in the analyses is 30 June 2017.

4. DESCRIPTION OF PARTICIPANT CHARACTERISTICS

4.1. Disposition

A flow of clusters (general practices) and participants through the trial will be summarised in a CONSORT diagram that will include the eligibility, reasons for exclusion, patients consenting, practices randomised to the two treatment groups, losses to follow up and the numbers analysed (for the primary outcome).

4.2. Baseline characteristics

The distributions of continuous variables will be examined. If data are approximately normally distributed, then the variable will be summarised in terms of the mean and standard deviation. For continuous variables that are not normally distributed median and inter quartile range (IQR) will be presented to summarise the variable. Categorical data will be summarised in terms of frequency counts and percentages. We will summarise all variables by trial arm, at both the cluster (practice) level and individual-level summary data. No formal statistical comparisons will be undertaken.

5. ASSESSMENT OF STUDY QUALITY

5.1. Eligibility checks

The numbers of patients who were eligible (identified as having multimorbidity), randomly selected, invited to participate (some excluded due to terminal illness for example), and agreeing to participate in the trial (and reasons for exclusion/no consent) will be reported in the CONSORT Flow Diagram. Amongst those eligible and invited to participate, socio-demographic and baseline characteristics will be compared between those consenting and those either refusing consent/not responding.

5.2. Study completion

Final follow up is at 15-month post-randomisation. The numbers of patients followed-up and lost to follow-up will be reported for each treatment arm in the CONSORT Flow Diagram.

5.3. Compliance/ Fidelity

Fidelity will be considered and examined in detail within the process evaluation (qualitative aspects of the process evaluation will be described elsewhere). Quantitative measures are detailed in Table 5 above and will be reported for the 3D intervention group only.

Compliance (at the patient level) will be defined as ‘full’ – two GP 3D appointments and two nurse 3D appointments attended; ‘partial’ – at least one GP or nurse 3D appointment attended, but not full attendance; and ‘none’ – no GP 3D appointment and no nurse 3D appointment attended. The percentage of patients in each of these categories will be reported for the 3D intervention arm.

5.4. Protocol deviations

Any protocol deviations will be fully documented.

6. ANALYSIS OF EFFECTIVENESS

The reporting and presentation of data from this trial will be in accordance with the CONSORT guidelines for cluster randomized trials[20]. STATA 14.1 will be used for all statistical analysis.

6.1. Summary of primary and secondary outcomes

The primary outcome will be summarised for each treatment group as the mean (SD) or median (IQR) as appropriate. Continuous secondary outcomes will be summarised as mean (SD) or median (IQR) as appropriate. Binary/ordinal data will be summarised in terms of frequency counts and percentages.

6.2. Primary analysis

The tested null hypothesis is that the mean quality of life (measured by the EQ5D-5L) for patients receiving the 3D intervention is the same as for those receiving usual care at 15 months follow up.

Primary analysis will take the form of mixed-effects multivariable linear regression, adjusted for practice (random effect to account for clustering), minimisation variables (practice size and practice deprivation score) and patient baseline EQ5D-5L. The results will be presented as the difference between group means, corresponding 95% confidence interval and P-value. The intra-class correlation (ICC) will also be reported, with a 95% confidence interval. Practices/patients will be analysed in the groups to which they were allocated and missing data will be imputed (section 3.3).

The distribution of EQ5D-5L and model residuals will be examined (a suitable transformation or a boot-strapped 95% confidence interval accounting for clustering[21] will be considered if necessary).

6.3. Secondary analyses

All analyses of secondary outcomes will be adjusted for the baseline measure of the outcome (if available) and minimisation variables. Patient level analyses will also adjust for practice (as a random effect). All secondary outcomes will be considered at 9 and 15 months.

The list of variables and planned models of analysis are given in the table below. Distributional checks will be carried out for all outcomes and the most appropriate models selected (for example, for ordinal outcomes mixed-effects ordered logistic models will be performed and assumptions regarding the ordinal nature of responses tested; if not valid alternative multinomial models will be employed). The effect, 95% confidence interval and P value will be reported for each model along with the ICC if possible.

Table 6: Planned analyses - Patient level secondary outcomes

Outcomes	Type of variable	Type of model[1]
CARE measure of relational empathy (GP)	Continuous	Mixed-effects linear regression
CARE measure of relational empathy (nurse)	Continuous	Mixed-effects linear regression
Care related to patients' priorities (LTC6)	Discrete (ordinal)	Mixed-effects ordered logistic regression
Care which is joined up (LTC6)	Discrete (ordinal)	Mixed-effects ordered logistic regression
PACIC measure of chronic disease management	Continuous	Mixed-effects linear regression
Overall satisfaction	Discrete (ordinal)	Mixed-effects ordered logistic regression
EQ5D-5L	Continuous	Mixed-effects linear regression
Self-rated health	Discrete (ordinal)	Mixed-effects ordered logistic regression
Bayliss measure of illness burden	Continuous	Mixed-effects linear regression
HADS Anxiety score	Continuous	Mixed-effects linear regression
HADS Depression score	Continuous	Mixed-effects linear regression
Multimorbidity Treatment Burden Questionnaire	Continuous	Mixed-effects linear regression
Medication adherence: MMAS-8	Continuous	Mixed-effects linear regression
Number of drugs prescribed	Count data	Poisson regression random effects
Number of high risk prescribing indicators	Count data	Poisson regression random effects
Continuity of care - COC index	Continuous	Mixed-effects linear regression
Continuity of care - Visit Entropy	Continuous	Mixed-effects linear regression
Quality of disease management	Continuous	Mixed-effects linear regression
Number of hospital admissions	Count data	Poisson regression random effects
Number of outpatient attendances	Count data	Poisson regression random effects

[1]Analyses will be dependent on distributional checks

Table 7: Planned analyses - Carer secondary outcomes

Carer secondary outcome	Type of variable	Type of model[1]
Carer experience scale	Continuous	Mixed-effects linear regression
EQ5D-5L carer score	Continuous	Mixed-effects linear regression
Multimorbidity treatment Burden Questionnaire (carers)	Continuous	Mixed-effects linear regression

[1]Analyses will be dependent on distributional checks

6.4. Process of care measures

Table 8: Planned analyses – Patient level process of care measures

Process measure	Type of variable	Type of model[1]
Number of primary care consultations with GP	Count data	Poisson regression random effects
Number of primary care consultations with nurse	Count data	Poisson regression random effects
Mean duration of face to face consultations in surgery with GP	Continuous	Mixed-effects linear regression
Mean duration of face to face consultations in surgery with nurse	Continuous	Mixed-effects linear regression
Number of different review consultations	Count data	Poisson regression random effects
At least one review received	Binary	Mixed-effects logistic regression

[1]Analyses will be dependent on distributional checks

Other process measures collected at 15 months in the intervention group only (reported in Table 5) will be reported with descriptive statistics. For several measures percentages will be reported both as the percentage of all patients recruited to the intervention arm and the percentage of patients receiving at least one nurse or GP review (as appropriate). See Table 9 below.

Table 9: Descriptive statistics - Patient level process of care measures collected in the intervention arm only

Process measure	Type of variable	Summary statistics	Denominator used in percentage		
			All patients in intervention arm	All those having at least one nurse review	All those having at least one GP review
Number of nurse 3D reviews	Ordinal	N (%)	✓	NA	NA
Number of GP reviews	Ordinal	N (%)	✓	NA	NA
Compliance[1]	Ordinal	N (%)	✓	NA	NA
Most important problem noted	Binary	N (%)	✓	✓	NA
EQ5D pain question noted	Binary	N (%)	NA	✓	NA
PHQ9 entered	Binary	N (%)	✓	✓	NA
Medication reviewed by Pharmacist (at least one comment entered)	Binary	N (%)	✓	NA	✓
Medication adherence noted	Binary	N (%)	✓	NA	✓
First patient goal noted	Binary	N (%)	✓	NA	✓
First plan noted (‘what patient can do’)	Binary	N (%)	✓	NA	✓

First plan noted ('what GP can do')	Binary	N (%)	✓	NA	✓
Patient agenda printed[1]	Binary	N (%)	✓[2]	✓[2]	NA
3D plan printed[1]	Binary	N (%)	✓[2]	NA	✓[2]
Number of times hospital physician was contacted	Count	Median (IQR)	NA	NA	NA

[1] Data not available in practices in Scotland

[2] Denominators include only those recruited from England

6.5. Sensitivity analysis

The analyses presented above make a number of assumptions about the data (for example, the treatment groups are balanced, all patients have received the allocated treatment, all patients completed questionnaires at the correct time point). Sensitivity analyses will consider whether the conclusions drawn in the primary analysis are sensitive or robust to different assumptions made.

6.5.1 Baseline imbalance

Baseline characteristics will be compared between the treatment groups and the magnitude of any differences considered in terms of their potential clinical importance (following discussion with clinicians). Where important differences exist, baseline characteristics will be adjusted for in sensitivity analyses.

6.5.2 Missing data

The impact of missing data on the primary analysis will be explored by conducting complete case analyses (using the complete cases data set (see section 3.1).

6.5.3 Non-compliance

If compliance is less than 90% we will also perform a complier average causal effect (CACE) analysis for the primary analysis of EQ5D-5L. Compliers can only be observed amongst those randomized to the 3D arm. Compliance will be defined as 'full' – two GP 3D appointments and two nurse 3D appointments attended; 'partial' – at least one GP or nurse 3D appointment attended but not full compliance; and 'none' – no GP 3D appointment and no nurse 3D appointment attended. These three categories will be compared in terms of key baseline characteristics. Sensitivity analysis considering the complier average causal effect (CACE), will include two analyses with a dichotomous indicator variable for compliance: one analysis will amalgamate patients in the 'full' and 'partial' groups; the other will combine those in the 'none' and 'partial' groups.

The CACE estimates will be obtained using instrumental variable regression including the same variables used in the primary analysis, randomized group as an instrumental variable and the indicator variable for compliance.

6.5.4 Time of questionnaire return

It is likely that there will be variation in the time at which questionnaires are completed. In sensitivity analyses, time of completion will be included as a covariate.

6.5.5 Outliers/influential values

The impact of outliers/influential values will be explored by removal of such observations from the analyses.

6.5.6 Treating deceased patients as missing for EQ5D-5L

As an additional sensitivity analysis, we will investigate the impact of treating the deceased as having a missing score for EQ5D-5L at 15 months rather than imputing a value of zero. This analysis will be performed using the complete case dataset.

6.6. Potential effect modifiers

Potential effect modifiers, selected *a priori* and informed by previous evidence, will be explored using appropriate interaction terms added to the regression models used for the primary analysis. Subgroup analyses of the primary outcome will explore differences in the effectiveness of the 3D intervention compared to usual care according to baseline measures of:

- Participant age (< median / ≥ median of consenting participants);
- Number of long term conditions as defined in section 2.4.3 (< median / ≥ median of consenting participants);
- Deprivation (quartiles of consenting participants);
- Depression alongside physical health problems (presence/absence).

Subgroup analyses are likely to be insufficiently powered since the trial was not powered to specifically test these effects. These subgroup analyses will therefore be hypothesis generating and will focus on interpretation of 95% confidence intervals rather than P-values.

6.7 Withdrawal rates

It is anticipated that a small number of patients will withdraw from each treatment group. Withdrawal refers to patients who actively decline further participation in the trial or are withdrawn by their practice (e.g. due to other illness, moving practice) and for whom there is no subsequent data collection. For each treatment group numbers and percentages of patients withdrawing will be reported along with reasons for withdrawal. Baseline characteristics will be compared between those who withdrew and those who did not in both treatment groups to look for differential withdrawal between the groups.

7. ANALYSIS OF SAFETY

In this population (older persons with multiple LTCs) a high number of adverse events (AE) are anticipated in both treatment groups, hence attention will be given only to serious adverse events (SAE) which may be related to the intervention or the research process (serious adverse reactions).

7.1. Serious adverse reactions

All reporting of SAEs of a related and unexpected nature will follow regulatory reporting requirements as set out in article 17 of the European Union Directive 2001. These will be reported to the sponsor immediately and will be reported to the REC within 7 days of the Trial Manager becoming aware of the event. Any relevant further information will be subsequently communicated within 8 days. In addition, all investigators will be notified. The TSC will be notified immediately of all SAEs thought to be treatment or research related. Potential SAEs which after review are not thought to be treatment or research related will be brought to the TSC's attention at their next scheduled meeting. The numbers and details of all SAEs will be reported to the Trial Management Group, Trial Steering Committee and Data Monitoring Committee.

For each treatment group the number and percentage of patients experiencing a serious adverse event, which appeared to be related to the intervention or the trial, will be reported. Given that patients with multimorbidity may be heavy users of secondary care services, new medical diagnoses, hospital admissions and deaths are expected and will not be considered as potential serious adverse events unless anyone involved in the study (participants, general practice staff or research staff) notify the research team of any events that they consider may have been related to the intervention or the research process. All deaths will be investigated for relatedness by requesting the patient's GP provide details of cause of death and relatedness to study.

If there are sufficient numbers of related SAEs, logistic regression with robust standard errors (to account for clustering) will be used to estimate the odds ratio (3D compared to usual care) for the different categories of SAEs. Corresponding 95% confidence intervals and P values will also be presented. If sufficient numbers of patients have multiple SAEs then ordered logistic models will be employed (e.g. outcome may be categorised as 0, 1, 2, 3+ SAEs).

7.1.1 Deaths

Given the population participating in the 3D trial, some deaths are expected before the end of follow-up in each treatment group. Numbers and percentages in each arm will be reported. Poisson regression and Cox regression with random effects will be considered (as appropriate) to calculate a rate ratio or hazard ratio with the corresponding 95% confidence intervals and P values. Models will also consider minimisation variables, age, number of long standing conditions, and EQ5D-5L at baseline.

8. CHANGES MADE TO STATISTICAL ANALYSIS PLAN

All amendments made to the statistical analysis plan (following approval by the TMG/ DMC and TSC of final version 1.0) will be listed in the table below. Following each amendment, a new version of the analysis plan will be created and previous versions saved.

Table 10: Record of amendments to statistical and economic analysis plans

Amendment	Rationale	Analysis plan version	Date SAP amended	Approved by (delete as appropriate)
				TMG / DMC / TSC
				TMG / DMC / TSC
				TMG / DMC / TSC
				TMG / DMC / TSC
				TMG / DMC / TSC
				TMG / DMC / TSC
				TMG / DMC / TSC

9. REFERENCES

1. Man MS, Chaplin K, Mann C, Bower P, Brookes S, Fitzpatrick B, Guthrie B, Shaw A, Hollinghurst S, Mercer S, Rafi I. Improving the management of multimorbidity in general practice: protocol of a cluster randomised controlled trial (The 3D Study). BMJ open. 2016 Apr 1;6(4):e011261.
2. Mann C, Shaw A, Guthrie B, Wye L, Man MS, Hollinghurst S, Brookes S, Bower P, Mercer S, Salisbury C. Protocol for a process evaluation of a cluster randomised controlled trial to improve management of multimorbidity in general practice: the 3D study. BMJ open. 2016 May 1;6(5):e011260.
3. Kind P, Hardman G, Macran S (1999) Population norms for EQ-5D. York: The University of York. Discussion paper 172 Discussion paper 172. 1-98 p.
4. Walters SJ, Brazier JE (2005) Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 14: 1523-1532.
5. Kennedy A, Bower P, Reeves D, et al. Implementation of self management support for long term conditions in routine primary care settings: cluster randomised controlled trial. BMJ 2013;346:f2882.
6. Nolan CM, Longworth L, Lord J, Canavan JL, Jones SE, Kon SS, Man WD. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax. 2016 Mar 30:thoraxjnl-2015.
7. McClure NS, Al Sayah F, Xie F, Luo N, Johnson JA. Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores. Value in Health. 2017 Jan 10.
8. https://www.ohe.org/publications/valuing-health-related-quality-life-eq-5d-5l-value-set-england
9. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Family practice. 2004 Dec 1;21(6):699-705.
10. Glasgow RE, Wagner EH, Schaefer J, Mahoney LD, Reid RJ, Greene SM. Development and validation of the patient assessment of chronic illness care (PACIC). Medical care. 2005 May 1;43(5):436-44.
11. Bayliss EA, Ellis JL, Steiner JF. Subjective assessments of comorbidity correlate with quality of life health outcomes: initial validation of a comorbidity assessment instrument. Health and Quality of life Outcomes. 2005 Sep 1;3(1):1.
12. Bocéréan C, & Dupret E. (2014). A validation study of the Hospital Anxiety and Depression Scale (HADS) in a large sample of French employees. BMC Psychiatry, 14, 354. http://doi.org/10.1186/s12888-014-0354-0
13. De las Cuevas C, Peñate W. Psychometric properties of the eight-item Morisky Medication Adherence Scale (MMAS-8) in a psychiatric outpatient setting. International Journal of Clinical and Health Psychology. 2015 Aug 31;15(2):121-9.
14. Guthrie B, McCowan C, Davey P, Simpson CR, Dreischulte T, Barnett K. High risk prescribing in primary care patients particularly vulnerable to adverse drug events: cross sectional population database analysis in Scottish general practice. BMJ. 2011 Jun 21;342:d3514.

15. Garrison GM, Bania B. Visit Entropy: comparing a novel method to existing continuity of care measures. European Journal for Person Centered Healthcare. 2015 Sep 3;3(3):343-51.

16. Bice TW, Boxerman SB. A quantitative measure of continuity of care. Medical care. 1977 Apr 1;15(4):347-9.

17. Al-Janabi H, Flynn TN, Coast J. Estimation of a preference-based carer experience scale. Medical Decision Making. 2011 May 1;31(3):458-68.

18. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality - An evaluation of different analytic approaches. Med Care. 2007 Jun;45(6):489-96.

19. Stevens JP. Outliers and influential data points in regression analysis. Psychological Bulletin. 1984 Mar;95(2):334.

20. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. Bmj. 2012 Sep 4;345:e5661.

21. Shiquan Ren , Hong Lai , Wenjing Tong , Mostafa Aminzadeh , Xuezhang Hou & Shenghan Lai (2010) Nonparametric bootstrapping for hierarchical data, Journal of Applied Statistics, 37:9, 1487-1498.