2020-12-22

The PARTNERS2 Study: Trial of Primary Care Based Collaborative Care for People with a Diagnosis of Schizophrenia, Bipolar or other types of Psychosis. Protocol version 1.0 dated 16.12.2020

Jones, Ben

http://dx.doi.org/10.24382/5028

Statistical & Health Economics Analyses Plan

Trial of Primary Care Based Collaborative Care for People with a Diagnosis of Schizophrenia, Bipolar or other types of Psychosis: The PARTNERS2 Study

ISRCTN 95702682

IRAS ID 13436

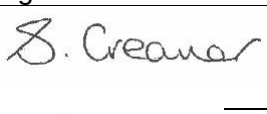
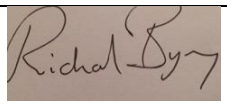
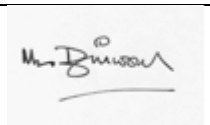
NIHR Programme Grant of Applied Research

(NIHR200625)

Contents

ADMINISTRATIVE INFORMATION

Title of Trial	Trial of Primary Care Based Collaborative Care for People with a Diagnosis of Schizophrenia, Bipolar or other types of Psychosis: The PARTNERS2 Study
Trial registration number	ISRCTN95702682
Protocol Version	7.3 (18.08.2020)
SAP Version	1.0 (16.12.2020)
SAP Revisions	Any revisions to the SAP after sign-off will be documented here, including a brief justification and timing of revision in relation to unblinding of data

	Name	Signature	Date
Statistical Analysis Plan Authored by:	Senior Statistician: Prof Siobhan Creanor		16/12/2020
	Trial Statistician: Mr Ben Jones		16/12/2020
	Trial Statistician: Dr Joanne Hosking		16/12/2020
	Health Economist: Prof Linda Davies		17/20/2020
Approved by:	Chief Investigator: Prof Richard Byng		17/20/2020
	Chief Investigator: Prof Max Birchwood		17/20/2020
	Independent Statistician: Dr Ben Carter		17/12/2020

ABBREVIATIONS

AE	Adverse Event
CACE	Complier Average Causal Effect
CBT	Cognitive Behavioural Therapy
COVID-19	Coronavirus Disease 2019
CSA	COVID-19 Sensitivity Analysis
CI	Confidence Interval
CRF	Case Report Form
EMA	European Medicines Agency
GP	General Practitioner
ICC	Intracluster Correlation Coefficient
ICER	Incremental Cost Effectiveness Ratio
ICECAP-A	ICEpop CAPability measure for Adults
IRR	Incidence Rate Ratio
ITT	Intention To Treat
MANSA	Manchester Short Assessment of Quality of Life
MAR	Missing at Random
MCAR	Missing Completely at Random
mITT	modified Intention To Treat
NB	Net Benefit
NICE	The National Institute for Health and Care Excellence
PP	Per Protocol
QALY	Quality Adjusted Life Year
QPR	Questionnaire about the Process of Recovery
RCT	Randomised Controlled Trial
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
TUS	Time Use Survey
(S)WEMWBS	(Short) Warwick-Edinburgh Mental Wellbeing Scale

1. INTRODUCTION

1.1. Background and rationale for the trial

The full background and rationale for PARTNERS2 is detailed in the study protocol [Version 7.3, 18.08.2020]. In summary, the PARTNERS2 study aims to examine the provision of collaborative care partners to better address the emotional, social and physical needs of people with severe mental illness in a co-ordinated way by placing a secondary care practitioner within general practice.

1.2. Purpose of statistical analysis plan

The study protocol includes an outline of the statistical methods to be employed in the analysis of the trial data. The purpose of the Statistical Analysis Plan (SAP) is to provide full details of the planned statistical methods to be used in the primary report of the trial results, in line with the Consolidated Standards of Reporting Trials (CONSORT) guidelines[1] and extensions for non-pharmacological trials[2] and cluster randomised trials[3], which provide guidelines to facilitate the complete and transparent reporting of clinical trials, as well as the ICH E9 Guidelines[4]. The PARTNERS2 protocol includes details of an internal pilot study within the definitive RCT. However, this SAP pertains only to the design and statistical analysis of the definitive trial. The planned health economic analyses are outlined in section 7. Details of the planned process evaluation, including detailed fidelity and adherence measures, are not discussed within this document.

2. TRIAL OBJECTIVES AND OUTCOME MEASURES

The overarching aim of this research is to establish the clinical and cost effectiveness of a primary care based model of collaborative care for adults with a clinical diagnosis of schizophrenia, bipolar, or other types of psychosis.

2.1. Primary objective

To assess whether provision of a care partner embedded within primary care can improve quality of life in adult patients with clinical diagnosis of schizophrenia, bipolar or other types of psychosis.

2.2. Primary outcome measure

The primary outcome measure is quality of life measured using version 2 (V2) of the Manchester Short Assessment of Quality of Life[5] (MANSA), which is detailed in Appendix 1.

The primary endpoint was originally planned for 10 months (± 1 month) post-randomisation (of GP practice/cluster). However, given a number of challenges with recruitment, followed by the start of the COVID-19 pandemic in early 2020, it was necessary to reduce follow-up time for the final participants recruited, scheduled to be followed-up in December 2020, to 9 months (± 1 month) post-randomisation, in order to keep within study timescales dictated by the funder.

2.3. Secondary outcome measures

- Number of hours per week spent in structured activity: Time Use Survey (TUS)[6–9]
- Personal Recovery: 15 item Questionnaire about the Process of Recovery (QPR-15)[10]
- Mental wellbeing: Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)[11] and the short version (SWEMWBS)[12]
- Experience of care (brief INSPIRE)[13]
- Capability (ICECAP-A) and Quality of Life (EQ-5D-5L)
- Healthcare monitoring:
    - Annual care check received
    - Blood pressure, weight, lipids, blood sugar, metabolic function assessed: checks recorded in primary care records and any interventions made.
    - Smoking, diet, alcohol: evidence in primary care records that weight has been recorded and evidence of advice/ lifestyle intervention offered
- Lifestyle outcomes: self-reported
    - Smoking (yes/no response and number of times smoking in a typical day)
    - Alcohol consumption (yes/no response and number of days drinking per week)
    - Cannabis use (yes/no response)

2.4. Safety outcomes

- Number of psychiatric hospital admissions
- Number of days inpatient as a result of psychiatric admission
- Crisis Care: Number of episodes under home treatment
- Crisis Care: Total days under home treatment

3. TRIAL DESIGN

3.1. General design

An open label, superiority, cluster randomised controlled trial in adult patients with a clinical diagnosis of schizophrenia, bipolar or other type of psychosis, with randomisation occurring at the GP practice level.

Patients aged 18 or over with a clinical diagnosis of schizophrenia, bipolar or other type of psychosis will be recruited from GP practices within one of three geographical regions in England: Birmingham, the South West (Devon and Cornwall) and Somerset. Practices with six or more potentially eligible patients and filtering in to one of the four localities (i) Birmingham and Solihull Mental Health NHS Foundation Trust; (ii) Livewell Southwest (Devon); (iii) Cornwall Partnership NHS Foundation Trust or (iv) Somerset Partnership NHS Foundation Trust will be eligible for participation in the study.

Once a practice has been recruited and authorised to begin recruitment of participants, invitations will be sent to potentially eligible patients. Once optimal recruitment within a practice has been reached (determined pragmatically according to numbers recruited, rate of recruitment and potential for further recruitment), and once all participants have consented to participate and completed the baseline assessment, the practice will be randomised to either continuation of usual care provision, or to receive the collaborative care intervention, and subsequently unmasked.

Participants within practices which have been randomised to the intervention group will receive the collaborative care intervention for up to twelve months, which includes a two month transition period back to usual care. Ten months (nine months for the final participants recruited, see section 2.2) (± 1 month) post-randomisation, participants will be invited to provide follow-up outcome data. Participants in control practices continue with treatment as usual (either involving secondary care or just within primary care as clinically required).

3.2. Blinding

Due to the nature of the intervention, neither the participants nor the assessors will be blinded to a participant's allocated group. The senior statistician and trial statisticians will remain fully blinded up until after database lock. The trial statisticians writing and running the analyses will remain pseudo-blinded to participants' allocated groups whilst undertaking the primary analyses.

3.3. Analysis populations

The primary analysis of the primary outcome will be based on the intention-to-treat (ITT) principle; participants will be analysed as randomised i.e. as their GP practice was randomised, regardless of their compliance with the trial protocol or lack of participation/completion if allocated to the intervention group. ITT is generally accepted to be the gold standard approach and provides a conservative estimate of the average intervention effect. The modified ITT (mITT) population for the primary analysis will include all participants with associated primary outcome data at both baseline and follow-up. The mITT will exclude only participants who were deemed ineligible following randomisation, those who withdrew from the trial and were unwilling for their previously collected data to be utilised or those who failed to provide both baseline and primary endpoint MANSA scores (i.e. there will be no imputation of missing primary endpoint MANSA scores for the primary analysis). Multiple imputation will be considered in order to achieve a full ITT population in sensitivity analyses (section 4.8).

One alternative to the ITT approach is a Per Protocol (PP) analysis; however, this approach could introduce bias into the trial through excluding participants after randomisation who did not participate/comply with the trial protocol, and thus jeopardising the group comparability achieved through randomisation. An alternative approach is complier-average causal effect (CACE) analysis, which provides an unbiased estimate of the intervention effect, based on participants who complied with their allocated group's protocol. The CACE analysis will be considered for the purposes of a sensitivity analysis, and as it is a less conservative approach than ITT, will likely result in a larger between-group difference. Two definitions of 'compliance' are provided in section 4.3. CACE analyses will be undertaken under certain conditions as outlined in section 6.6.1.

Additional Per Protocol (PP) sensitivity analyses are planned to explore the impact of: (i) follow-ups being undertaken outside of the pre-specified primary endpoint window (section 6.6.2) and (ii) care partner availability within the intervention period (section 6.6.3).

Additional sensitivity analyses are required in order to explore the effect of the COVID-19 outbreak, which began during the trial. Details are provided in section 6.9.

The pre-specified safety outcomes (section 6.11) and the reported serious adverse events will be analysed on an as-treated basis, with participants who had at least one documented interaction with a care partner categorised as being in the intervention group.

3.4. Inclusion and exclusion criteria

GP practices

Recruited practices will:

- Filter into: (i) Birmingham and Solihull Mental Health NHS Foundation Trust; (ii) Livewell Southwest; (iii) Cornwall Partnership NHS Foundation Trust or (iv) Somerset Partnership NHS Foundation Trust;
- Have at least six potentially eligible patients.

Working with GP practices is a key feature of the PARTNERS2 intervention. GP practices were screened for capacity and capability to deliver the trial. This included: willingness to meet and physically host the Care Partner (including computer access); the ability to meet data collection requirements; no participation in any competing studies and at least 16 patients on their SMI register.

Individuals

The study population will comprise patients who:

- are registered with a GP practice which satisfies the practice-level inclusion criteria;
- are aged 18 or over;
- have a clinical diagnosis of schizophrenia, bipolar or other type of psychosis; and
- have evidence for care need in relation to their diagnosis in the previous two years (automatic for those in secondary care, and assessed from patient notes for primary care only).

Potential participants will be excluded if they:

- have an inability to understand English (and lack of access to translation services);
- have an inability to give informed consent;
- have more significant need requiring ongoing secondary multi-disciplinary care (such as those meeting criteria for diagnostic cluster 13, assertive outreach or early intervention functions);
- are currently receiving home crisis care or are in an inpatient or secure setting;
- are excluded at the discretion of GPs, if it is felt that inclusion in the trial is not in the best interest of the patient;
- are currently participating in a Cognitive Behavioural Therapy (CBT), psychosocial or medicinal trial for psychosis or bipolar;
- have a primary diagnosis of dementia receiving secondary care for dementia;

- have a primary diagnosis of learning disability receiving care from secondary care for learning disability;
- have an ongoing significant and chaotic substance or alcohol misuse making engagement with the trial and intervention problematic.

4. STATISTICAL PRINCIPLES

4.1. Randomisation, stratification and allocation concealment

As a cluster randomised controlled trial, randomisation will be undertaken at the GP practice level. Practices will be randomised on a 1:1 basis to provide trial participants with either the collaborative care intervention, or usual care.

Practices will be stratified by locality (Birmingham and Solihull Mental Health Foundation Trust, Livewell Southwest (Devon), Cornwall Partnership NHS Foundation Trust, and Somerset Partnership NHS Foundation Trust) and practice size (large versus small). Practice size will be determined by the number of adults in the practice registered on the practice Quality and Outcomes Framework register, classified specifically under MH001 (adults with schizophrenia, bipolar affective disorder or other psychoses and other patients on lithium therapy). A 'large' practice will be one in which there are 70 or more MH001 patients, and a 'small' practice one with less than 70 MH001 participants. Allocation will be achieved by means of minimisation[14] with no random element.

To ensure allocation concealment from participants, clinicians and data collectors, once a practice within a locality has been recruited, eligible patients within each practice are approached and only following completion of participant recruitment and participant-level baseline assessments, is the allocation of the practice unmasked. The randomisation procedure will be undertaken by a member of the Peninsula Clinical Trials Unit (PenCTU) who is not involved in participant assessments or delivery of the intervention or usual care.

4.2. Sample size calculation

The sample size for the full trial is based on detecting a mean difference of 0.45 points in the overall MANSA score. Assuming a standard deviation (SD) of 0.9, this is equivalent to detecting half a standard deviation. This standardised effect size of 0.5 corresponds to a medium to large effect size, the same as was pre-specified in the sample size calculation for the DIALOG+ trial[15].

To detect the target difference of 0.45 points, and assuming a standard deviation (SD) of 0.9, coefficient of variation of cluster size of 0.74 and intra-cluster correlation of 0.05[16] the recruitment target for the full trial was originally 336 participants across ~56 GP clusters. This assumed a mean of 6 participants recruited per cluster and 20% drop out at the individual participant level and 10% drop out at the cluster level following randomisation.

The trial protocol allowed for an interim blinded review of the assumptions underpinning the original sample size calculation and, in particular, facilitated the exploration of an adjustment for the correlation between baseline and 10 month MANSA scores. This pre-specified review was undertaken using data from the first 39 participants with complete baseline and follow-up primary outcome data and explored the *a priori* assumption of a correlation of 0.5 between baseline and follow-up MANSA scores. The point estimate of this correlation was 0.69 (80% confidence interval: 0.56 to 0.79) and as such, it was deemed appropriate to conservatively allow for a correlation of 0.5 in a revised sample size calculation. Retaining the other underpinning assumptions of the original sample size calculation indicates a revised recruitment target of 270 participants from ~45 GP practices, to achieve 90% power or 204 participants from ~34 GP practices to achieve 80% power, to detect the pre-specified between-group difference of 0.45 points.

In December 2019, both the Data Monitoring and Trial Steering Committees approved a revised recruitment target of 270 participants.

In late January 2020, the study funder mandated that trial recruitment be terminated at the end of February 2020, regardless of the revised recruitment target. By mid-January 2020, 170 participants had been recruited to the trial. Following the funder's decision, the aim is to recruit 204 participants, from ~34 GP practices, to achieve 80% power to detect the pre-specified target difference of 0.45 points, based on the underpinning assumptions of SD of 0.9, mean cluster size of 6 participants, coefficient of variation of 0.74, intra-cluster correlation of 0.05 and correlation between baseline and follow-up MANSA scores of 0.5 and allowing for 20% drop out at the individual participant level and 10% drop out at the cluster level following randomisation.

4.3. Intervention fidelity and protocol deviations

4.3.1. Fidelity to the intervention

Failure to follow the intervention protocol amongst intervention participants may occur either through lack of interest or engagement from the participant with their care partner, or from a lack of care partner availability or a lack of adherence by the care partner to the protocol. While the intervention is designed to be flexible, significant deviations from delivery as intended need to be documented. Appropriate summary statistics will be used to describe the level of participant engagement with the intervention and care partner availability.

Two definitions of intervention fidelity have been constructed, which are believed to represent a minimum level of intervention engagement from which a benefit could be expected:

i. A participant has attended at least six sessions (face-to-face, by telephone or virtually) with a care partner, with each session lasting a minimum of ten minutes;
ii. A participant has attended at least four sessions with their care partner in which goals were discussed.

4.3.2. Follow-up outside of pre-specified visit window

The numbers and proportions of participants with follow-up undertaken outside of the pre-specified window (10 months post-randomisation ± 1 month, except for the final 56 participants recruited, whose pre-specified window is 9 months ± 1 month, as per protocol v7.3,18.08.2020) will be summarised by allocated group. Length of follow-up will also be described.

Any other protocol deviations will be documented and reported to the Chief Investigator and Sponsor.

4.4. Planned interim review to assess sample size assumptions

A planned blinded interim review of the underpinning sample size assumptions has been undertaken, as outlined in section 4.2.

4.5. Collection of outcome measures

Outcome measures will be collected from participants at baseline (prior to randomisation), and again at the primary endpoint (9 or 10 months (±1 month) post-randomisation).

Table 1: Outcome measures

	Baseline	Primary endpoint
Primary Outcome		
MANSA	x	x
Secondary Outcomes		
TUS	x	x
QPR-15	x	x
WEMWBS/SWEMWBS	x	x
Brief INSPIRE	x	x
ICECAP-A	x	x
EQ-5D-5L	x	x
Healthcare Monitoring		x
Lifestyle outcomes	x	x
Safety variables		
Admissions (psychiatric)		x
Crisis care (home treatment)		x

4.6. Time points of statistical analysis

Comparative statistical analysis, with the exception of any unplanned interim analysis requested by the oversight committees or Funder, will be undertaken after the final follow-up data has been collected for each participant and the database locked.

4.7. Data sources and data quality

The data will come from information entered onto Case Report Forms (CRFs) completed at baseline and 10 months (or 9 months for the last 56 participants recruited). A ± 1 month tolerance will be allowed for collection of outcome measures, after which point a protocol deviation will be recorded. At each visit, participants will be given the option to complete the self-report questionnaires alone or alongside the researcher. Following collection, data will be entered onto a secure RedCap Cloud database.

4.8. Missing data

4.8.1. Missing outcome data

The intention is to use complete case data for the primary analysis of the primary outcome, on a modified intention-to-treat (mITT) basis, where any participants for whom an overall MANSA score cannot be calculated at either/both baseline and follow-up will be excluded from the primary analysis, which is of the change in overall MANSA score. Participants who missed some items within the MANSA will be included in the mITT population, provided few enough items were missed to allow valid imputation of the overall score, as detailed in section 4.9. The sample size calculation allowed for up to 20% loss to follow-up at the individual level, and 10% loss to follow-up at the cluster level, mitigating the risk of loss of power as a result of incomplete data, although there is a risk of bias if there is differential loss to follow-up between the intervention and control groups, or if loss to follow-up is not random within allocated groups.

A sensitivity analysis to assess the effect of missing data on the primary outcome will be undertaken. Multiple imputation will be undertaken via a joint modelling approach using multivariate linear regression. More detail of the proposed procedure can be found in section 6.6.4.

Similarly, analyses of secondary outcomes will be undertaken on a mITT basis, with only participants who provided the relevant data at both baseline assessment and follow-up included. Where applicable, these analyses will include participants who missed some items within the outcome under consideration, provided few enough were missed to allow valid imputation of the total score. Imputation of items within each outcome will be considered as appropriate in line with published guidelines, with further details for each outcome specified in section 4.9.

4.8.2. Other missing data

Other missing demographic data such as sex and age will be queried following data entry, although it is not expected that there will be a considerable amount of such missing data.

4.9. Derived outcome variables

In order to ensure the accuracy of the results, the calculation of each of the derived outcomes will be programmed twice, independently by two statisticians.

Overall MANSA score (primary outcome): The MANSA V2 questionnaire is included in Appendix A. The overall score is calculated as the mean of the eleven domain-specific questions, excluding the ‘life as a whole’ questions (Q1 & Q25). The overall score is the mean of the following:

- Job: (Q7a OR Q7b)
- Financial Situation (Q9)
- Leisure: (Q10)
- Social Life: (Q13 AND Q14)
- Living situation: (Q16 AND Q18a OR Q18b)
- Family: Q20
- Safety: Q22
- Health: (Q23 AND Q24)

In any cases where participants erroneously answer both 'or' questions (i.e. Q7 & Q18), the value to include in the calculation of the overall score will be determined by referring to supporting evidence within the questionnaire. Specifically, if a participant answers both Q7a and Q7b, Q7a will be used if Q4 indicates the participant is in work, and Q7b will be used if Q4 indicates a participant is not in work. Similarly, Q18a will be used if Q17 indicates that a participant lives with other people, and Q18b will be used if Q17 indicates that the participant lives alone. If it is not possible to determine the appropriate choice of question, the participant's score for that question will be deemed as missing.

Where between one and five of the answers to the eleven questions are missing, the overall MANSA score will be calculated as the mean of the non-missing items. Similar to guidance in a published manual for the 12 item version of the MANSA[17] (in which it was suggested that the average score could be calculated provided no more than six of the twelve items were missing), if more than five items are missing, this overall MANSA score will be deemed missing. The 11 item version used in PARTNERS2 was first used in a study undertaken in 2009[18]. Overall MANSA scores can range from 1-7, with higher scores representing a better quality of life.

Time Use Survey[6–9]**:** The TUS comprises questions on seven components of time use: the number of hours spent in (i) paid work; (ii) looking for work; (iii) education and training; (iv) voluntary work; (v) leisure activities; (vi) sporting activities and (vii) socialising. The outcome of interest is the total mean number of hours per week spent in structured activity over the past month. Within the CRF, for each component, the total number of hours per month is collected. For each of the monthly total number of hours collected, the weekly average will be calculated by multiplying by twelve, then dividing by 52. The final value of interest, the mean number of hours per week spent in structured activity over the past month, will be calculated as the sum of the weekly mean of each of the seven components.

Process of Recovery (QPR-15)[10]**:** This questionnaire comprises 15 questions, each answered on a 0-4 scale (0 = disagree strongly, 1 = disagree, 2 = neither agree nor disagree, 3 = agree, 4 = agree strongly). The overall score is the total of each of the 15 items and will therefore fall between 0-60, where higher scores indicate a more positive experience of recovery. As there are no published guidelines on imputation of missing items, any missing items will be imputed as the mean of the completed items, provided no more than three items are missing. The overall score will be classified as missing if more than three items are missing.

Warwick Edinburgh Mental Wellbeing Scale (WEMWBS)[11]**:** The WEMWBS comprises 14 questions, each answered on a 1-5 scale (1 = none of the time, 2 = rarely, 3 = some of the time, 4 = often, 5 = all of the time). The overall score is the total of each of the 14 items and will therefore fall between 14-70, where higher scores represent a greater degree of wellbeing. As proposed in the WEMWBS user guide[19], imputation of missing items will be undertaken by using the mean of the completed items, provided the number of missing items is no greater than three. The overall score will be classified as missing if more than three items are missing.

Short Warwick Edinburgh Mental Wellbeing Scale (SWEMWBS)[12]**:** The SWEMWBS comprises a subset of seven (Q1-3; Q6-7; Q9 and Q11) of the 14 items included in the WEMWBS. The final score is calculated as the sum of these seven items, which is then converted according to Table 2 below:

Table 2: Raw score conversion table for SWEMWBS

Raw Score	Metric Score	Raw Score	Metric Score
7	7.00	22	19.98
8	9.51	23	20.73
9	11.25	24	21.54
10	12.4	25	22.35
11	13.33	26	23.21
12	14.08	27	24.11
13	14.75	28	25.03
14	15.32	29	26.02
15	15.84	30	27.03
16	16.36	31	28.13
17	16.88	32	29.31
18	17.43	33	30.70
19	17.98	34	32.55
20	18.59	35	35.00
21	19.25		

Brief INSPIRE[13]**:** This questionnaire comprises five questions, each answered on a 0-4 scale (0 = not at all, 1 = not much, 2 = somewhat, 3 = quite a lot, 4 = very much). The overall score is the sum of the five scores, multiplied by 5 to give a score ranging from 0-100, where higher scores indicate a more positive experience of support. The overall score will not be calculated if answers to any of the five questions are missing.

EQ-5D-5L: This will be estimated from the published EQ-5D-5L Crosswalk Index Value Calculator, as recommended NICE at the time of analysis[21,22]. The EQ-5D-5L responses for each domain are scored on a scale of 1 to 5, where 1 represents no problems and 5 represents extreme problems[23]. These scores are then combined to generate a health profile for each person. For example, if a participant rates themselves as having no problems on each domain, the profile is 11111; if the participant reports no problems with mobility (1), some problems with self-care (2), some problems with usual activity (2), moderate pain or discomfort (3) and extremely anxious or depressed (5), their profile would be 12235.

The UK utility value set contains a population-based utility weight for each possible profile[22,24]. In this example, the utility for a person reporting no problems on each of the dimensions is 1. The utility weight for the person with a profile of 12235 is 0.176. Death has a utility of 0. Some profiles have a negative utility. For example, if the participant with a profile of 12235 reported extreme rather than moderate pain and discomfort, their profile would be 12255 and the relevant utility value is -0.088.

ICECAP-A: This questionnaire comprises five questions, in which each response within each question has a tariff value attached, as shown in Table 3, where level 4 represents full capability and level 1 represents no capability. The overall score is calculated as the sum of each of the five relevant tariff scores[20] as shown in Table 3 below:

Table 3: Tariff scores for responses to each ICECAP-A response

1. Feeling settled and secure	
Level 4	0.222
Level 3	0.191
Level 2	0.101
Level 1	-0.001
2. Love, friendship and support	
Level 4	0.228
Level 3	0.189
Level 2	0.096
Level 1	-0.024
3. Being independent	
Level 4	0.188
Level 3	0.156
Level 2	0.084
Level 1	0.006
4. Achievement and progress	
Level 4	0.181
Level 3	0.159
Level 2	0.091
Level 1	0.021
5. Enjoyment and pleasure	
Level 4	0.181
Level 3	0.154
Level 2	0.069
Level 1	-0.003

Healthcare Monitoring: The healthcare monitoring data comprises fourteen items pertaining to whether or not a participant has received various surveillance and active management services through a primary care notes review. The surveillance items include an annual health check, a blood pressure recording, a weight recording, a check of lipids, and a blood glucose check. For each participant, the number of healthcare checks received will be calculated to give a total between 0-5. For the active management items, it is not appropriate to calculate an aggregate as the provision of active management is dependent on the clinical need for that management, which may or may not have been documented in the participant's records. As such, no derivation is required for the active management items.

Safety Data: The number of psychiatric admissions, inpatient days as a result of psychiatric admission, crisis care events and days under home care will be obtained from a notes review after the primary endpoint follow-up.

5. STUDY POPULATION

Data from the screening process through to the completion of the trial will be recorded and presented in a CONSORT flow diagram, for both individual participants and GP practices. In particular, the following data will be provided:

- Number of practices screened for eligibility
- Number of practices excluded (ineligible or declined)*
- Number of eligible practices
- Number of individuals assessed for eligibility and approached
- Number of individuals excluded (ineligible or declined)*
- Number of participants at baseline assessment
- Number of practices randomised
- Number of practices randomised to each allocated group
- Number of participants randomised to each allocated group
- Number of practices who received their allocated treatment
- Number of practices who did not receive their allocated
- Number of practices lost to follow-up and withdrawn, with reasons*
- Number of additional participants lost to follow-up/withdrawn*
- Number of practices analysed
- Number of participants analysed

*Reasons will be provided where available

5.1. Participants and practices who discontinue, withdraw or are lost to follow-up

It is possible that participants or entire GP practices (clusters) may withdraw consent part way through the trial, or an individual's allocated intervention may be discontinued on medical grounds.

If a GP practice withdraws after randomisation, any patients registered at the practice who have consented to the trial will be withdrawn, and participant data already collected will be retained unless explicit request is made to destroy it. If a practice withdraws after randomisation, consideration will be given to replacing with another practice, although the sample size calculation has allowed for cluster-level dropout of 10%. In this scenario, any baseline data collected will be summarised if appropriate and if not jeopardising the anonymity of participants.

Participants who discontinue will be categorised into one of the following:

- Continue to consent for follow-up and data collection
- Consent to use pre-collected data only
- Withdrawn consent to use any data

Participants who withdraw from the study, or whose intervention is discontinued on medical grounds, will not be replaced although their available data will be used unless they have specifically requested for it to be removed from the database.

Reasons for withdrawal or loss to follow-up at both the individual and cluster-level will be summarised in the CONSORT diagram where available.

5.2. Baseline characteristics and demographics

Baseline characteristics at both the cluster and individual level, collected prior to randomisation, will be cross-tabulated according to allocated group to check for balance between groups and provide an overview of the study sample. The variables at the individual level will include the outcome measures outlined in section **Error! Reference source not found.**, as well as demographic information, and at the practice level will include practice size and locality (the stratification factors).

Summary statistics for continuous measures will be reported as means, standard deviations and ranges where the distribution appears normal, and as medians, inter-quartile ranges and ranges if the distribution is skewed. Categorical data will be summarised by frequencies and percentages. Formal statistical comparison between randomised groups at baseline is not good practice[25,26] and thus will not be performed; relative balance between allocated groups is expected at both the individual and cluster level. Any considerable imbalance will be noted, assessed for relevance and, where appropriate, additional adjustments will be considered in secondary sensitivity analyses.

6. STATISTICAL ANALYSES

6.1. General considerations

Wherever possible, analyses will be presented with 95% confidence intervals and all reported p-values will be two-sided, unless otherwise stated, with hypothesis testing undertaken at the 5% significance level. As there is a single primary outcome, no adjustment for multiple testing will be undertaken.

For continuous variables, summary information will be presented in the form of means (including mean difference between follow-up and baseline where appropriate) alongside standard deviations and ranges for outcomes that are normally distributed. Summary information for non-normally distributed continuous outcomes will be presented in the form of medians, inter-quartile ranges and ranges. Binary and ordinal variables will be presented as frequencies and percentages. If there is evidence that the distributional assumptions of the modelling approaches employed are violated, suitable variable transformations will be explored, and the use of bootstrapped confidence intervals will be considered.

6.2. Adjustments

Primary analyses of the outcomes will be adjusted for the cluster-level stratification variables as fixed effects (hereafter referred to as fully adjusted models), including locality (Birmingham, Devon, Cornwall and Somerset), practice size (large vs small), and (individual level) baseline measure, where available, in line with European Medicines Agency (EMA) guidance on adjustment for baseline covariates[27]. For completeness, the primary analyses of the outcomes will be repeated without adjustment for stratification variables, but with adjustment for baseline measure as a fixed effect, where available (hereafter referred to as partially adjusted models).

Both the fully and partially adjusted models will include GP practice as a random effect in order to account for the potential correlation amongst participants within the same practice, as is recommended for the analysis of CRCTs[28]. As the statistical modelling will be undertaken at the individual level, the denominator degrees of freedom will reflect this. A Kenward-Roger adjustment of the degrees of freedom will be undertaken for each of the models.

6.3. Intra-cluster correlation coefficients

Intra-cluster correlation coefficients (ICCs) are useful not only for describing the extent of clustering within the trial data, but also for providing other researchers with evidence for which to design future studies. As a result, the ICC will be presented for each outcome for both the fully and partially adjusted models (as defined in section 6.2). ICCs derived from crude models (i.e adjusted only for allocated group) where the dependent variable is the outcome measured at the primary endpoint (rather than the change from baseline) will also be presented for each outcome. All ICCs will be presented alongside 95% confidence intervals calculated using Stata's `estat icc` command, which uses the delta method to estimate the standard error of the ICC[29].

6.4. Primary analysis of the primary outcome

The primary outcome is the Manchester Short Assessment of Quality of Life (MANSA) V2 overall score (questionnaire included in Appendix A), measured at follow-up as well as at baseline. The questionnaire comprises a total of 25 questions, and the primary outcome will be derived as the mean of the eleven domain-specific questions (further details in section 4.9).

Descriptive summary statistics (e.g. means and standard deviations) will be presented for the primary outcome of MANSA scores at baseline and follow-up by allocated group, and overall. Change in overall MANSA score will be calculated for each participant by subtracting the score at baseline from the score at follow-up. The primary analysis will compare the change in overall MANSA scores between the two allocated groups. Although the sample size calculation was presented under a framework in which the MANSA score at the primary endpoint was to be analysed, (as opposed to the change in MANSA between baseline and primary endpoint) both approaches yield identical results, provided that baseline adjustment is made[27]. As a result, the choice to model the change is based solely on interpretability. Analyses will be undertaken using a Gaussian random effects regression model, including the cluster-level stratification variables (locality and practice size) and individual-level baseline MANSA score as fixed effects covariates, and GP practice as a random effect. Utilising the baseline MANSA scores within the model will increase the precision of the estimated intervention effect.

Partially adjusted analyses (defined in section 6.2), will also be presented for completeness. Both fully and partially adjusted between-group comparisons will be presented with 95% confidence intervals, p-values and ICCs.

The primary analysis of the primary outcome will be conducted twice, independently, but as much of the analysis as capacity allows will be double-coded.

6.5. Interpretation of primary analysis results

The primary analysis will test the following null and alternative hypotheses:

H_0: There is no difference in the change in MANSA score from baseline to primary endpoint follow-up assessment between the two treatment groups.

H_1: There is a difference in the change in MANSA score from baseline to primary endpoint follow-up assessment between the two treatment groups.

If the results of the primary adjusted analyses suggest that there is sufficient evidence to reject the null hypothesis, it will be concluded that the collaborative care intervention is significantly different to the standard treatment currently available in improving quality of life. However, whilst the study is designed around the hypothesis testing framework outlined above, emphasis will be placed on appropriate interpretation of the 95% confidence intervals.

6.6. Secondary analyses of the primary outcome

6.6.1. Intervention fidelity sensitivity analysis

Frequencies and percentages for each of the two fidelity definitions outlined in section 4.3 will be presented. If ≥20% of participants in the intervention group with valid baseline and follow-up primary outcome data are classified as non-compliers **and** if ≥ 20% are classed as compliers according to the fidelity definitions in section 4.3.1, a sensitivity analysis based on this definition will be triggered.

One potential method of sensitivity analysis to explore the effect of intervention fidelity is a Per Protocol (PP) analysis, where only participants receiving a pre-specified minimum level of allocated treatment are analysed. However, this approach has the potential to introduce bias as a result of excluding participants and thus jeopardising the between-group comparability achieved through randomisation.

As an alternative approach, a complier-average causal effect (CACE) analysis of the primary outcome will be undertaken for each of the two fidelity definitions if triggered according to the above criteria. This will provide an unbiased estimate of the intervention effect after taking account of intervention receipt. In a CACE analysis, a participant's allocated group is included in the model as an instrumental variable, and the participant's *given* treatment is an endogenous variable, which in this case will consist of either the collaborative care intervention, or usual care. A participant's given treatment will be determined according to the fidelity definitions outlined in section 4.3. Under fidelity definition (i), the given treatment of a participant allocated to the intervention arm would remain as intervention provided they attended at least six ten-minute sessions with a care partner, and would otherwise become usual care. Under fidelity definition (ii), the given treatment of a participant allocated to the intervention arm would remain as intervention provided they discussed goals in at least four sessions with their care partner, and would otherwise become usual care. Under both definitions, the given treatment of participants allocated to the usual care arm would always remain as usual care, as it is not possible for control group participants to receive the collaborative care intervention. In order to account for the clustering inherent in data collected from a CRCT, instrumental variable regression with standard errors that are robust to clustering will be implemented.

The results of the outlined CACE analyses will be presented alongside 95% confidence intervals, and interpreted in the context of their robustness to the primary analysis of the primary outcome.

6.6.2. Follow-up window sensitivity analysis

An additional PP sensitivity analysis of the primary outcome will be undertaken, in order to account for any potential impact of participants being followed up outside their pre-specified primary endpoint window; participants followed up early may not yet have had ample opportunity to benefit from the intervention, and any intervention effect amongst participants followed up late may have lessened or disappeared. As such, for the primary outcome only, the fully adjusted and the partially adjusted analysis models will be fitted to the subset of the population who provided follow-up data within the pre-specified window.

6.6.3. Care partner availability sensitivity analysis

An additional PP analysis of the primary outcome will be undertaken, in order to account for any potential impact of care partner availability during the intervention period of the trial. Specifically, any practices randomised to the intervention arm in which a care partner was available for less than 70% of the first 10 months of intervention period (from time of unmasking until planned primary endpoint) will be excluded from this sensitivity analysis.

6.6.4. Missing data sensitivity analysis

A sensitivity analysis to assess the effect of missing data on the primary outcome will be attempted. Multiple imputation will be undertaken via a joint modelling approach using multivariate linear regression and including random effects at GP practice (cluster) level[30], as well as all variables included in the fully adjusted model (see section 6.2). Additional auxiliary variables not used in the main analysis model (e.g. number of CP sessions) may also be included if found to increase the performance of the imputation.

A Markov Chain Monte Carlo procedure will be used to generate a total of 20 imputed datasets, with a suitable burn-in period and interval between each imputed dataset. Convergence of the chains will be assessed visually (using autocorrelation and traceplots) and using the $\hat{R}$ statistic. The primary analysis model will be fitted to each of the imputed datasets and the results combined using Rubin's rules[31]. The multiple imputation and associated analysis will be undertaken in R using the `pan` and `mitml` packages[32], which allow for clustering within the multiple imputation model.

6.7. Further exploratory analysis of the primary outcome

6.7.1. Additional MANSA questions

In addition to the primary outcome of overall MANSA score, calculated as described in section 4.9, it is possible to obtain additional information from the MANSA questionnaire. Specifically:

Life in general

- Feelings about life on the whole today, measured on a Likert scale from 1 (terrible) to 7 (delighted)
- Feelings about life as a whole, measured on a Likert scale from 1 (terrible) to 7 (delighted)

Life opportunities

- Areas of participants' life that they wished to improve on, but weren't able to

Work and education

- Months in full/part time work over the past two years
- Employment Status
- Average hours of work per week

Finance

- Frequency of difficulty meeting household bills

Social

- Friendship (two yes/no responses)

Living situation

- Type of residence
- Cohabitees

Family

- Frequency of contact with relatives

Safety

- Victim of violence (yes/no response)

Each of these will be summarised appropriately at baseline and at follow-up by allocated group, with between-group differences in changes from baseline and 95% confidence intervals (CIs) presented where appropriate.

6.7.2. Planned subgroup analyses

Whilst likely underpowered, some additional exploratory subgroup analyses will be undertaken, and will be limited to the primary outcome and the following subgroups:

- Diagnostic groups (Bipolar vs Schizophrenia or other psychosis)
- Locality (Birmingham vs Devon vs Cornwall vs Somerset)
- Care provider (primary care vs secondary care) based on usual point of care (at time of screening)
- Practice Size (small (<70 participants on the Quality and Outcomes Framework (QOF) register) vs large (≥70 participants on the QOF register)

The effects of these subgroups will be explored by including the treatment group by subgroup interaction in separate regression models including adjustment for baseline MANSA score, stratification factors and the relevant subgroup term as well as including GP practice as a random effect. The results of these analyses will be treated with caution and interpreted as exploratory[33], rather than for the purposes of definitive hypothesis testing. As a result, the interaction terms will be presented alongside 95% CIs. Associated p-values will also be presented but interpretation will focus on the 95% CIs. Graphical presentation of these results will be considered in order to aid interpretation.

6.8. Analysis of secondary outcomes

Secondary analyses will involve considering the change at the primary endpoint from baseline for each of the secondary outcome measurements.

As with the primary analysis, continuous secondary outcomes (TUS, QPR-15, Brief INSPIRE, WEMWBS and SWEMWBS, EQ-5D-5L and ICECAP-A) will be analysed using Gaussian random effects regression models with adjustment for the cluster-level stratification variables (locality and practice size) and individual-level baseline measurement where available as fixed effects, and with adjustment for GP practice as a random effect. The fully adjusted mean differences between the two groups will be presented alongside 95% confidence intervals and p-values. In addition, partially adjusted analyses in which adjustment is made only for the baseline value of the outcome (as a fixed effect) and GP practice (as a random effect) will be presented with 95% confidence intervals for the mean difference and p-values. ICCs will be presented for all fully and partially adjusted models.

For the surveillance element of the healthcare monitoring data, the total number of healthcare checks received will be calculated between 0-5. This data will be modelled using a fully and partially adjusted Poisson regression model. The incidence rate ratio (IRR) between allocated groups will be reported alongside 95% CIs and the associated ICC.

For the active management services, it is not appropriate to interpret the totals in the same manner as for the surveillance services, as the provision of active management is dependent on the clinical need for that management, which may or may not have been documented in the participant's records. As a result, for each of these items, the number of participants in receipt will be presented alongside percentages and 95% CIs, by allocated group and overall, but no formal inferential analysis will be undertaken.

For lifestyle outcomes (smoking, alcohol consumption and cannabis use), no inferential analysis will be undertaken. Instead, the number and percentage of participants who self-report as smokers, drinkers and cannabis users will be summarised by allocated group at baseline and follow-up by allocated group and overall. Amongst smokers, appropriate summary statistics of the number of times smoking per day will also be reported, as will the average number of days per week drinking amongst those who self-report drinking alcohol.

The distribution of each of the secondary outcomes and the model assumptions will be examined visually and, where necessary, appropriate transformations will be sought. In any cases where transformations of the outcome are unsuccessful in satisfying the modelling assumptions, alternative methods of analysis will be considered and explored. This may include bootstrapped confidence intervals if appropriate.

Details of the planned health economic analysis are in section 7 below.

6.9. COVID-19

Due to social distancing measures introduced in response to the COVID-19 outbreak in early 2020, government guidelines led to a restriction on face-to-face interaction with trial participants, impacting both the delivery of the care partner intervention and/or usual care, and the collection of outcomes measures. As a result, it was necessary to (i) amend delivery of the intervention to facilitate contact with trial participants via telephone or videoconference facilities and (ii) conduct follow-up outcome collection visits remotely. An 8-week trial period was undertaken in order to determine the feasibility of delivering these elements of the trial remotely.

The EMA guidance on the implications of COVID-19 on methodological aspects of ongoing trials[34] recommends careful consideration of the possible impact of the COVID-19 pandemic on trial delivery and the integrity of the results. The PARTNERS2 trial had completed GP practice and participant recruitment, baseline data collection and randomisation by the time social distancing measures were introduced on 23/03/2020. However, 29 practices still had primary endpoint CRFs and participant questionnaire follow-ups to complete, with 139 participant follow-ups still to be done, and with the final participants due to provide follow-up data in December 2020. During the COVID-19 lockdown, both care partner interactions and outcome data collection were undertaken remotely, either via the telephone, or via videoconferencing facilities. Although social distancing restrictions were later relaxed somewhat, a decision was taken to continue outcome data collection remotely, due to ongoing uncertainties surrounding the introduction of further restrictions.

An additional questionnaire was designed in order to capture the effects of the COVID-19 pandemic on various aspects of the trial participants' lives, including but not limited to their mental and physical health. This questionnaire collects both qualitative and quantitative data. Handling of the qualitative data is beyond the scope of this SAP, but the quantitative data will be summarised appropriately by allocated group, and overall, although no formal inferential analysis will be undertaken.

The COVID-19 pandemic and associated social distancing measures have the potential to impact the results of the trial in two ways: firstly, through impacting the nature of the collaborative care intervention, which originally was developed to be delivered predominantly through face-to-face interaction; and secondly, by its potential to directly influence outcome measures (e.g. the TUS, and elements of the MANSA pertaining to living situation, social and leisure activities, etc).

In order to aid the assessment of the impact of the pandemic on the trial data, participants will be categorised as either 'pre COVID-19', or 'during COVID-19', with any participants providing follow-up data after 23/03/2020 categorised as the latter. After discussion with the Trial Management Group, it was agreed that, in addition to the primary outcome, the TUS was particularly likely to be affected by the lockdown measures. As such, three additional COVID-19 Sensitivity Analyses (CSA) will be undertaken for both the MANSA and the TUS:

> **CSA1**: Refitting of the fully and partially adjusted primary analysis models, excluding participants providing follow-up data after 23/03/2020. Interpretation will focus on comparison against the treatment effect estimate obtained from the mITT primary analysis.

CSA2: Refitting of the fully and partially adjusted primary analysis models, with an additional treatment x COVID-19 categorisation interaction term. Interpretation will focus on the point estimate and 95% CI of this interaction term.

CSA3: Refitting of the fully and partially adjusted primary analysis models, with an additional adjustment for the binary COVID-19 categorisation covariate. Interpretation will focus on the point estimate and 95% CI of this adjustment, and will aid in the exploration of the effect of the lockdown on the outcome measures (rather than the impact of COVID-19 on the estimated treatment effect).

This SAP will be reviewed regularly throughout the pandemic and updated as required in line with the latest developments, and taking cognisance of any updated methodological guidance.

6.10. Model assumptions

The assumptions underpinning the statistical models will be visually assessed (e.g. using residual versus fitted plots and QQ plots), with suitable transformations explored as necessary if these assumptions are not met. If there is a suggestion of substantial violation of the model assumptions, firstly suitable transformation of the outcome will be sought. If unsuccessful, bootstrapped confidence intervals for the treatment effect estimates will be considered, as will analyses of cluster-level aggregates. The presence of outliers will also be visually assessed and additional sensitivity analyses considered if appropriate.

6.11. Safety outcomes

The safety outcomes, total number psychiatric admissions and days inpatient as a result of psychiatric admission, and the total number of crisis care episodes and days under home treatment as a result of a crisis care episode, are collected at the primary endpoint via a secondary care notes review. These will be descriptively summarised on an as treated basis (at least one interaction with a care partner) and overall alongside 95% CIs where appropriate; no formal hypothesis testing of the safety outcomes will be undertaken.

6.12. Serious adverse events

The risk of harm associated with the trial procedures and the intervention are considered to be low, and therefore no non-serious adverse events will be collected for this trial.

Serious adverse events (SAEs) will be reported from the date of consent up until 12 months after the point of unmasking the practice (i.e. for the equivalent of the entire intervention period including the final two month transition period back to usual care for participants in practices allocated to the intervention group). Further details of the reporting procedures for SAEs can be found in the trial protocol. The numbers and percentages of participants with one or more reported SAEs, as well as the numbers and percentages of reported SAEs, will be presented on an as treated (at least one interaction with a care partner) basis. Summary statistics will also be presented of SAE classification, assessment of relatedness to trial treatment (not/unlikely/possibly/probably/definite related) and categorisation of severity (mild/moderate/severe) of reported SAEs. No formal statistical testing will be undertaken of the SAE data.

6.13. Statistical software

The statistical analyses will be undertaken using StataSE version 14 or later and R.

7. Health Economics Analysis

7.1. Overall approach

The aim of the economic evaluation is to estimate the cost-effectiveness of a primary-care based model of collaborative care for adults with a clinical diagnosis of schizophrenia, bipolar, or other types of psychosis (intervention) compared to usual practice (comparator).

The economic analysis will use patient-level data collected at baseline and during follow-up. The perspective of the primary analysis is NHS and Social Care (costs) and people with a clinical diagnosis of schizophrenia, bipolar, or other types of psychosis (health benefits).

The primary objectives of this analysis are to:

- Estimate the costs of health and social care in the intervention and comparator groups, and assess whether there are differences between groups;
- Estimate the QALYs gained by patients in the intervention and comparator groups, and assess whether there are differences between groups; and
- Assess whether any additional benefit is worth any additional cost.

7.2. Time horizon

The time horizon of the primary analysis will be 10 months (nine months for the final participant recruited, see section 2.2), the period from baseline to end of scheduled follow-up (± 1 month).

7.3. Intervention and comparator

The analysis will compare the collaborative care intervention to usual practice. The trial protocol [Version 7.3, 18.08.2020] describes the intervention and usual care in detail.

7.4. Primary outcome

The primary outcome of the economic analysis is the incremental cost effectiveness ratio (ICER), which combines service use costs and health benefit by dividing the difference between interventions in costs (net costs) by the difference in health benefit (net health benefit):

$$\text{ICER} = \frac{\text{Cost}_{\text{intervention}} - \text{Cost}_{\text{usual practice}}}{\text{Health benefit}_{\text{intervention}} - \text{Health benefit}_{\text{usual practice}}}$$

7.5. Primary measure of health benefit

The measure of health benefit for the primary economic analysis will be Quality-Adjusted Life Years (QALY) estimated from the EQ-5D-5L. The EQ-5D-5L is a validated, generic health status measure, designed to compare health outcomes across diseases. NICE recommends the QALY and the EQ-5D-5L as measures for economic evaluations[35].

The EQ-5D-5L captures five domains of health status: mobility, self-care, usual activity, pain/distress, and anxiety/depression. Each domain uses a five-point scale: no problems, slight problems, some problems, severe problems, or unable to do activity. The summary measure for the EQ-5D in the economic analysis is utility. This will be estimated from the published EQ-5D-5L Crosswalk Index Value Calculator, as recommended NICE at the time of analysis and outlined in more detail in section 4.9.

QALYs will be estimated as:

$$QALY = \Sigma[(U_i + U_{i+1}) /2] \times (t_{i+1} - t_i) \text{ over } i = 0 \text{ \& } 1$$

where U = utility and t = time at assessment. The time between assessments is the time from baseline to primary endpoint follow-up (i = 1).

Sensitivity analysis will assess whether alternative measures of health benefit could change the conclusions of the economic analysis. These alternatives are first, the ICECAP-A as an alternative measure to generate utility values and QALY's. The summary measure for the ICECAP-A is a tariff value (utility). This will be estimated from the published guidance[20]. The primary clinical outcome will also be used as an effectiveness measure of health benefit. For the purposes of the economic evaluation this will be defined as the change in score from baseline to follow up.

7.6. Resource use and cost estimation

The direct costs of services use will be estimated from the reported service use multiplied by the unit cost of that service type, for each participant. Published national unit costs will be used (NHS reference costs database[36] and the Unit Costs of Health and Social Care[37]). The price year for all costs will be that of the most recent published unit costs.by summing the cost of each resource used to provide health and social care (expected to be 2019-20).

The sources of service use data to estimate costs will be:

(i) Case note review for the safety outcomes (number of psychiatric hospital admissions, days in hospital and number of episodes of home-based crisis care, days of home-based crisis care), GP, practice nurse and other GP-practice consultations (mental and physical health). The case note review will cover the period from baseline to end of scheduled follow-up.

(ii) Interview with participants at baseline and end of scheduled follow up to collect information about their use of other primary, secondary and community health care and social care services not covered in the case note review. This will record service use over the last 3 months.

Collection of service use data from case notes across the multiple care providers involved in providing services (primary, secondary and community health care and social care) is complex, highly resource intensive and costly in the absence of integrated record linkage for individual service users. Accordingly, within the trial budget constraints, case note data collection was focussed on key services that could be expected to differ if the intervention was effective. The rationale for choosing a 3-month, rather than 10-month, recall period for participant reported service use data included: concerns about the burden to participants of recalling service use over a longer period and the need to balance complete service use data with incomplete recall, potentially high levels of inconsistent or missing data, limited resources for data collection.

Regression analysis will be used to estimate a cost per day for participants with complete 3-month service use data collected at the follow up interview. The regression model will include participants' baseline costs, the costs of psychiatric hospital admission and home-based crisis care and GP/practice nurse/other health professional consultations at the GP practice over the follow up period. The estimated cost per day will be used to impute the cost of time between baseline to start of 3-month recall period at follow up that is not spent as a psychiatric hospital inpatient. Sensitivity analysis will explore whether alternative methods of estimating 10-month costs from the 3-month data change the conclusions of the economic analysis.

Costs for participants with incomplete 3-month service use data or incomplete data from case note review will be estimated using the approach to dealing with missing observations or follow up described in section 7.7.

7.7. Missing data

The cost and health benefit data will be analysed by treatment allocated and include data for all participants whether or not they completed planned care. However missing data are inevitable from loss to follow-up or missing observations. Single imputation will be used for missing baseline measures of cost, utility and clinical indicators[38], but not missing demographic data. An indicator for missing demographic data will be used.

The final imputation strategy will depend on the pattern of missingness in the data. If the observed data approximate MAR or MCAR (including missingness dependent on observed covariates), multiple imputation from available data will be used, as recommended by Faria et al[39]. Alternatives to multiple imputation for the primary analysis include complete case analysis, indicator analysis or single imputations. However, these are planned for sensitivity analyses unless there is evidence that multiple imputation is inappropriate (missing not at random, given identified dependence on observed covariates). The sensitivity analyses will assess the robustness of the results to assumptions about missing data (see Table 4).

Literature review and regression analysis of pooled data (masked to treatment allocation) will be used to identify key baseline and follow-up variables associated with costs and QALYs to include in the imputation models. The data will be imputed by category of cost and EQ-5D domain to make best use of available data.

7.8. Primary analysis

The EQ-5D-5L and ICECAP-A data will be summarised as proportions of participants reporting each level (n/N, %, 95% confidence interval). Descriptive analyses of costs and summary health benefit measures, such as utility or preference weights (mean, standard deviation, 95% confidence interval) will be reported for the intervention and comparator for the complete case and imputed data at baseline and primary endpoint. Costs will be reported for the 3-month and primary endpoint cost estimates by inpatient admission, home-based crisis care and other health and social care services. The latter will be reported by primary, secondary and community health care and social care.

Regression analysis will be used to estimate net costs and QALYs of the intervention. Prior to analysis, the pooled data (masked to treatment allocation) will be used to identify the appropriate distribution and regression model for costs and QALYs. Covariates to account for factors that influence costs or QALYs will be identified from published literature, supplemented with analysis of pooled (masked) baseline data. Baseline covariates are expected to include: age; sex; ethnicity; service use costs before baseline; utility at baseline; time since diagnosis; living arrangements (e.g. alone, with partner).

The trial cluster (GP practice) and stratification variables will be included (locality and practice size) as defined for the statistical analysis of clinical outcomes.

The regression-based estimates of costs and outcomes will be bootstrapped to replicate 10,000 pairs of incremental cost and QALY outcomes of the intervention; The distribution of pairs of net costs and QALYs will be plotted on the cost effectiveness plane[35].

ICERs estimate the net cost per QALY gained by an intervention, and raise the question whether that cost is worth paying. To address this, the ICERs are compared with how much decision makers may be willing to pay for an additional QALY. However, the UK has no universally agreed cost-effectiveness threshold. The National Institute for Health and Care Excellence[40] in the past suggested a threshold of £20,000 to £30,000 per QALY. However, other commentators suggest this may be lower[41,42]. Reflecting this lack of consensus, the monetary value of simulated QALYs will be estimated across the range of £0 to £30,000 willingness to pay thresholds. This recognises that decision makers may not be willing to pay for an additional QALY (in other words they may only seek the lowest cost option) or could be willing to pay up to £30,000 for an extra QALY gained. To estimate the likelihood the intervention is cost effective for the primary analysis, a willingness to pay threshold of £15,000 (the mid-point of the £0 to £30,000 range) will be used.

The monetary values of the bootstrapped net QALY estimates will be used to estimate the net benefit statistic (NB) from each pair of simulated net costs and QALYs as:

NB statistic = (QALY x threshold willingness to pay value) – Cost.

Cost effectiveness acceptability curves (CEAC) will be generated to show whether the likelihood that the intervention is cost-effective at each willingness to pay value is above 50%. As decision makers increase what they are willing to pay for an extra QALY, the additional health benefits from an intervention become more valuable, and it achieves net benefit in a bigger proportion of the 10,000 replicates.

7.9. Sensitivity analysis

Table 4 shows the planned sensitivity analyses to assess whether and how the study design affects the conclusions of the economic evaluation.

7.10. Statistical software

STATA SE version 14 or later will be used to conduct imputation with predictive mean matching and sequential chained equations as well as all regression and bootstrapping analyses. SPSS version 23 will be used for preliminary costing of service use data.

Table 4: Sensitivity analyses for health economics analyses

Focus	Change	Rationale
Measure of health benefit	**From EQ-5D-5L to:** MANSA	Sensitivity analysis will explore cost-effectiveness of the intervention using the primary clinical measure of effectiveness rather than the generic QALY health benefit measure. Criterion is whether the incremental cost-effectiveness ratio switches from the primary analysis.
	ICECAP-A	Sensitivity analyses will explore cost-effectiveness of the intervention using a preference-based measure of capability, rather than the EQ-5D-5L, to estimate QALYs. The criterion will be the likelihood that the cost per QALY gained is cost effective.
Estimating 10-month costs from 3-month participant reported service use data for services other than hospital admissions and home-based crisis care	**From single imputation to:** Assume constant monthly cost of services collected for a 3-month recall period and estimate full 10-month cost prop-rata. For services collected for a 3-month recall period, assume the intervention increases their use in the 3 months following baseline as participants are signposted to and engage with services	The full 10-month costs will be estimated from different sources of data collected for different time periods. Using alternative methods allows an assessment of the impact of methods to account for this on the results and the robustness of the primary analysis.
Accounting for missing data due to loss to follow up or missing items	**From multiple imputation to (or method used for primary analysis):** Complete case analysis Indicator analysis Single imputation	The validity and robustness of techniques to account for missing observations and loss to follow up will decrease as the extent of missing data increases. Using alternative methods allows an assessment of the impact of methods to account for missing data on the results and the robustness of the primary analysis.

REFERENCES

1. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:869. doi:10.1136/bmj.c869

2. Barbour V, Bhui K, Chescheir N, et al. CONSORT Statement for randomized Trials of nonpharmacologic treatments: A 2017 update and a CONSORT extension for nonpharmacologic Trial Abstracts. *Ann Intern Med*. 2017;167(1):40-47. doi:10.7326/M17-0046

3. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: Extension to cluster randomised trials. *BMJ*. 2012;345(7881). doi:10.1136/bmj.e5661

4. EMA. *ICH Topic E 9 Statistical Principles for Clinical Trials Step 5 NOTE FOR GUIDANCE ON STATISTICAL PRINCIPLES FOR CLINICAL TRIALS*.; 1998. Accessed March 12, 2020. http://www.emea.eu.int

5. Priebe S, Huxley P, Knight S, Evans S. Application and results of the Manchester Short Assessment of Quality of Life (MANSA). *Int J Soc Psychiatry*. 1999;45(1):7-12. doi:10.1177/002076409904500102

6. Fowler D, Hodgekins J, French P, et al. Social recovery therapy in combination with early intervention services for enhancement of social recovery in patients with first-episode psychosis (SUPEREDEN3): a single-blind, randomised controlled trial. *The Lancet Psychiatry*. 2018;5(1):41-50. doi:10.1016/S2215-0366(17)30476-5

7. Hodgekins J, Birchwood M, Christopher R, et al. Investigating trajectories of social recovery in individuals with first-episode psychosis: A latent class growth analysis. *Br J Psychiatry*. 2015;207(6):536-543. doi:10.1192/bjp.bp.114.153486

8. Hodgekins J, French P, Birchwood M, et al. Comparing time use in individuals at different stages of psychosis and a non-clinical comparison group. *Schizophr Res*. 2015;161(2-3):188-193. doi:10.1016/j.schres.2014.12.011

9. Gershuny J. *Time-Use Surveys and the Measurement of National Well-Being* .; 2011.

10. Law H, Neil ST, Dunn G, Morrison AP. Psychometric properties of the Questionnaire about the Process of Recovery (QPR). *Schizophr Res*. 2014;156(2-3):184-189. doi:10.1016/j.schres.2014.04.011

11. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Dinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health Qual Life Outcomes*. 2007;5. doi:10.1186/1477-7525-5-63

12. Ng Fat L, Scholes S, Boniface S, Mindell J, Stewart-Brown S. Evaluating and establishing national norms for mental wellbeing using the short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. *Qual Life Res*. 2017;26(5):1129-1144. doi:10.1007/s11136-016-1454-8

13. Williams J, Leamy M, Bird V, et al. Development and evaluation of the INSPIRE measure of staff support for personal recovery. *Soc Psychiatry Psychiatr Epidemiol*. 2015;50(5):777-786. doi:10.1007/s00127-014-0983-0

14. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ*. 2005;330(7495):843. doi:10.1136/bmj.330.7495.843

15. Priebe S, Golden E, Kingdon D, et al. Effective patient–clinician interaction to improve treatment outcomes for patients with psychosis: a mixed-methods design. *Program Grants Appl Res*. 2017;5(6):1-160. doi:10.3310/pgfar05060

16. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res*. 2000;9(2):81-94. doi:10.1177/096228020000900202

17. Van Nieuwenhuizen C, Janssen-de Ruijter EA., Nugter M. *Manual: Manchester Short Assessment of Quality of Life (MANSA)*.; 2017. Accessed October 21, 2020. https://www.researchgate.net/publication/321881972_Manual_Manchester_Short_Assessment_of_Quality_of_Life_MANSA

18. Evans S, Huxley PJ, Maxwell N, Huxley KLS. System-level change in mental health services in North Wales: An observational study using systems thinking. *Int J Soc Psychiatry*. 2014;60(4):337-351. doi:10.1177/0020764013489672

19. Stewart-Brown S, Janmohamed K. *Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) User Guide*.; 2008. Accessed May 15, 2020. http://www.mentalhealthpromotion.net/resources/user-guide.pdf

20. Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Heal Econ (United Kingdom)*. 2015;24(3):258-269. doi:10.1002/hec.3014

21. Position statement on use of the EQ-5D-5L value set for England (updated October 2019). Accessed November 10, 2020. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l

22. Van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Heal*. 2012;15(5):708-715. doi:10.1016/j.jval.2012.02.008

23. EuroQol Research Foundation. *EQ-5D-5L User Guide*. https://euroqol.org/publications/user-guides

24. Dolan P. Modeling Valuations for EuroQol Health States. *Med Care*. 1997;35(11):1095-1108. doi:10.1097/00005650-199711000-00002

25. Senn S. Testing for baseline balance in clinical trials. *Stat Med*. 1994;13(17):1715-1726. doi:10.1002/sim.4780131703

26. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:869. doi:10.1136/bmj.c869

27. European Medicines Agency. *Guideline on Adjustment for Baseline Covariates in Clinical Trials*.; 2015. Accessed April 24, 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf

28. Eldridge SM, Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research.* John Wiley & Sons; 2012.

29. StataCorp. estat icc — Estimate intraclass correlations. Accessed May 15, 2020. https://www.stata.com/manuals/meestaticc.pdf

30. Taljaard M, Donner A, Klar N. Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biometrical J*. 2008;50(3):329-345.

doi:10.1002/bimj.200710423

31. Rubin DB, Wiley J, York N, Brisbane C, Singapore T. *Multiple Imputation for Nonresponse in Surveys*.; 1987.

32. Grund S, Lüdtke O, Robitzsch A. Multiple Imputation of Multilevel Missing Data: An Introduction to the R Package pan. *SAGE Open*. 2016;6(4):215824401666822. doi:10.1177/2158244016668220

33. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine - Reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357(21):2189. doi:10.1056/NEJMsr077003

34. European Medicines Agency. *Points to Consider on Implications of Coronavirus Disease (COVID-19) on Methodological Aspects of Ongoing Clinical Trials*. Accessed October 5, 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-implications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical_en-0.pdf

35. *Guide to the Methods of Technology Appraisal 2013*.; 2013. Accessed November 10, 2020. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781

36. National Cost Collection for the NHS. Accessed November 24, 2020. https://www.england.nhs.uk/national-cost-collection/

37. Unit Costs of Health and Social Care 2019. Accessed November 24, 2020. https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2019/

38. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med*. 2005;24(7):993-1007. doi:10.1002/sim.1981

39. Faria R, Gomes M, Epstein D, White IR. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. *Pharmacoeconomics*. 2014;32(12):1157-1170. doi:10.1007/s40273-014-0193-3

40. NICE. Carrying NICE over the threshold. Published 2020. Accessed November 10, 2020. https://www.nice.org.uk/news/blog/carrying-nice-over-the-threshold

41. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and care excellence cost-effectiveness threshold. *Health Technol Assess (Rockv)*. 2015;19(14):1-503. doi:10.3310/hta19140

42. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: What it is and what that means. *Pharmacoeconomics*. 2008;26(9):733-744. doi:10.2165/00019053-200826090-00004

APPENDIX A: MANSA V2 questionnaire

MANSA QUALITY OF LIFE ASSESSMENT – V2

LIFE IN GENERAL

1. HOW DO YOU FEEL ABOUT YOUR LIFE AS A WHOLE, TODAY? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

LIFE OPPORTUNITIES

2. IN THE PAST YEAR, HAVE THERE BEEN TIMES WHEN YOU WANTED TO IMPROVE ANY OF THE FOLLOWING ASPECTS OF YOUR LIFE BUT WERE UNABLE TO? (Please tick all boxes that apply)

Work or education	☐	Living situation	☐
Finances	☐	Family life	☐
Leisure	☐	Safety	☐
Social life	☐	Health	☐

WORK AND EDUCATION

3. HOW MANY MONTHS HAVE YOU WORKED (PART-TIME OR FULL-TIME) IN THE PAST 2 YEARS? ☐☐ *months*

4. WHAT IS YOUR CURRENT EMPLOYMENT STATUS? (Please tick one box only)

In paid work	☐	Not working – looking after the home	☐
In sheltered work	☐	Unemployed	☐
In training / education	☐	Retired	☐
Not working – due to long-term illness or disability	☐	Other	☐

5. **If working:**
HAVE YOU WORKED CONTINUOUSLY OVER THE PAST 3 MONTHS? *Yes* ☐ *No* ☐

6. **If working:**
ON AVERAGE, HOW MANY HOURS A WEEK DO YOU WORK? ☐☐ *Hours*

7a. **If working:**
HOW DO YOU FEEL ABOUT YOUR JOB? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

7b. **If not working:**
HOW DO YOU FEEL ABOUT NOT WORKING? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

FINANCE

8. HOW FREQUENTLY (IF AT ALL) DO YOU FIND IT DIFFICULT TO MEET THE COST OF HOUSEHOLD BILLS? (Please tick one box)

All of the time	Most of the time	Some of the time	Seldom	Never
☐	☐	☐	☐	☐

9. HOW DO YOU FEEL ABOUT YOUR FINANCIAL SITUATION? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

LEISURE

10. HOW DO YOU FEEL ABOUT YOUR LEISURE ACTIVITIES? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

SOCIAL

11. DO YOU HAVE ANYONE WHO YOU WOULD CALL A 'CLOSE FRIEND'? Yes ☐ No ☐

12. IN THE PAST WEEK HAVE YOU SEEN A FRIEND? Yes ☐ No ☐

13. HOW DO YOU FEEL ABOUT THE NUMBER OF FRIENDS YOU HAVE? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

14. HOW DO YOU FEEL ABOUT THE QUALITY OF YOUR FRIENDSHIP(S)? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

LIVING SITUATION

15. IN WHICH TYPE OF RESIDENCE DO YOU CURRENTLY LIVE? (Please tick one box only)

House or flat (owned)	☐	Sheltered housing	☐
House or flat (rented)	☐	Residential home	☐
Boarding out (inc B&B)	☐	Nursing home	☐
Mobile Home	☐	Hospital ward	☐
Hostel/ Supported or Group Home	☐	Homeless	☐

16. HOW DO YOU FEEL ABOUT YOUR ACCOMMODATION? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

17. WHO DO YOU LIVE WITH (IF ANYBODY) IN YOUR CURRENT HOME? (Please tick all boxes that apply)

Live alone	☐	Children over 18	☐
Spouse / partner	☐	Other family	☐
Parent(s)	☐	Non-family	☐
Children under 18	☐		

LIVING SITUATION CONTINUED

18a. **If living with other people:**

HOW DO YOU FEEL ABOUT THE PEOPLE THAT YOU LIVE WITH? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

18b. **If living alone:**

HOW DO YOU FEEL ABOUT LIVING ALONE? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

FAMILY

19. HOW OFTEN DO YOU HAVE CONTACT WITH A RELATIVE (NOT INCLUDING THOSE WHO LIVE WITH YOU) EITHER FACE TO FACE OR BY TELEPHONE? (Please tick one box only)

- Not at all ☐
- Daily ☐
- At least weekly ☐
- At least monthly ☐
- At least 3 monthly ☐
- At least yearly ☐
- Less than yearly ☐

20. HOW DO YOU FEEL ABOUT YOUR RELATIONSHIP WITH YOUR FAMILY? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

SAFETY

21. IN THE PAST YEAR, HAVE YOU BEEN A VICTIM OF VIOLENCE? Yes ☐ No ☐

22. HOW DO YOU FEEL ABOUT YOUR PERSONAL SAFETY? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

HEALTH

23. HOW DO YOU FEEL ABOUT YOUR HEALTH? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

24. HOW DO YOU FEEL ABOUT YOUR PRESENT MENTAL HEALTH? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

LIFE OVERALL

25. HOW DO YOU FEEL ABOUT YOUR LIFE AS A WHOLE? (Please mark one box only)

Terrible		Mostly dissatisfied		Mostly satisfied		Delighted
1	2	3	4	5	6	7
	Displeased		Mixed		Pleased	

Thank you for completing the MANSA

APPENDIX B: Examples of figures and tables for the primary publication

Table B1: Baseline demographic data (cluster level)

Outcome	Intervention (M=)	Control (M=)	Overall (M=)
N (%) Locality			
- Birmingham - Cornwall - Devon - Somerset			
N (%) Practice Size			
- Large - Small			
Mean (SD) Median [Range] Practice Size			

M = number of practices

Table B2: Baseline demographic data (individual level)

Outcome	Intervention (N=)	Control (N=)	Overall (N=)
N (%) Gender			
- Female - Male - Non Binary - Prefer not to say			
N (%) Transgender			
- Yes - No - Questioning/unsure - Prefer not to say			
N (%) Ethnicity			
- White - Mixed ethnicity - Asian - Black - Other ethnic group - Not known/not provided			
N (%) Relationship Status			
- Single - Married - Civil Partnership - Partner - Divorced/Separated - Widowed - Other - Prefer not to say			
N (%) Education Level			
- Level 1-2 - Level 3-5 - Level 6-8 - None - Other			

N (%) Physical Health			
- Chronic Heart Disease - Cancer - Stroke - Bronchitis/COPD/ emphysema - Asthma - Diabetes - Epilepsy, seizures or fits - Hypertension - Liver disease - Kidney disease			
N (%) Smoker			
- Cigarettes - Cigars - Pipe - E-cigarettes			
N; mean (SD) [range] times smoking per day			
N (%) Drinker			
N; median (IQR) times drinking per week			
N; mean (SD) [range] age at mental health diagnosis			
N (%) Mental Health medication			
- Antipsychotics - Antidepressants - Sedatives/ Hypnotics - Mood stabilisers - Other			
N (%) with formal carer			
N (%) with informal carer			
N (%) with childcare responsibilities			
N; mean (SD) [range] hours of childcare in the past month			
N (%) with other caring responsibilities			
N; mean (SD) [range] hours of other caring responsibilities in the past month			
N (%) with household tasks			
N; mean (SD) [range] hours of household tasks per week			

Table B3: Intervention compliance

Mean (SD) [Range]	Intervention
Number of GP practices where a care partner was in place for at least part of the trial period	
Number of GP practices where a care partner was in place for all of the trial period	
Number of interactions with care partner	
Number of interactions with care partner lasting at least 10 minutes	
Number of interactions with care partner with goals discussed	

Table B4: Primary outcome analysis - MANSA

N; Mean (SD) [Range]	Intervention (N =)			Control (N =)			Partially Adjusted[1] Treatment Effect[3] (Intervention – Control) [95% CI] p-value	Fully Adjusted[2] Treatment Effect[3] (Intervention – Control) [95% CI] p-value
	Baseline MANSA	Month 10 MANSA	Change	Baseline MANSA	Month 10 MANSA	Change		
Primary Analysis (mITT)								
Sensitivity Analysis 1 (PP – follow-up window)								
Sensitivity Analysis 2 (PP – care partner availability)								
Sensitivity Analysis 3 (Multiple Imputation)								

1 Adjusted for baseline MANSA and including random effects for GP practice

2 Adjusted for baseline MANSA as well as the stratification variables (practice size and locality) and including random effects for GP practice

3 Treatment effect is the between-group difference in the change in MANSA between baseline and primary endpoint

Table B5: Primary outcome analysis – complier average causal effect (CACE)

N; Mean (SD) [Range]	Compliers			Control & Non-Compliers			Partially Adjusted[1] Treatment Effect[3] (Intervention – Control) [95% CI] p-value	Fully Adjusted[2] Treatment Effect[3] (Intervention – Control) [95% CI] p-value
	Baseline MANSA	Month 10 MANSA	Change	Baseline MANSA	Month 10 MANSA	Change		
CACE (i)								
CACE (ii)								

1 Adjusted for baseline MANSA and including random effects for GP practice

2 Adjusted for baseline MANSA as well as the stratification variables (practice size and locality) and including random effects for GP practice

3 Treatment effect is the between-group difference in the change in MANSA between baseline and primary endpoint

Table B6: Primary outcome – additional summary statistics

N; Mean (SD) [Range] Median [IQR]	Intervention (N =)			Control (N =)		
	Baseline	Month 10	Change	Baseline	Month 10	Change
Life as a whole, today						
Life as a whole						
N (%) Life Opportunities						
- Work/Education			*NA*			*NA*
- Finances			*NA*			*NA*
- Leisure			*NA*			*NA*
- Social Life			*NA*			*NA*
- Living Situation			*NA*			*NA*
- Family Life			*NA*			*NA*
- Safety			*NA*			*NA*
- Health			*NA*			*NA*
Months in work in the past 2 years						
Hours working per week						
N (%) Employment Status						
- Paid work			*NA*			*NA*
- Sheltered work			*NA*			*NA*
- Training/education			*NA*			*NA*
- Not working - illness/disability			*NA*			*NA*
- Not working – looking after the home			*NA*			*NA*
- Unemployed			*NA*			*NA*
- Retired			*NA*			*NA*
- Other			*NA*			*NA*
N (%) Difficulty meeting household bills						
- All of the time			*NA*			*NA*
- Most of the time			*NA*			*NA*
- Some of the time			*NA*			*NA*
- Seldom			*NA*			*NA*
- Never			*NA*			*NA*
N (%) close friend			*NA*			*NA*
N (%) seen a friend in the past week			*NA*			*NA*
N (%) frequency of contact with a relative						
- Not at all			*NA*			*NA*
- Daily			*NA*			*NA*
- At least weekly			*NA*			*NA*
- At least monthly			*NA*			*NA*
- At least 3 monthly			*NA*			*NA*
- At least yearly			*NA*			*NA*

- Less than yearly	*NA*	*NA*
N (%) Living Situation		
- House/flat (owned)	*NA*	*NA*
- House/flat (rented)	*NA*	*NA*
- Boarding out	*NA*	*NA*
- Mobile Home	*NA*	*NA*
- Hostel/supported/group home	*NA*	*NA*
- Sheltered housing	*NA*	*NA*
- Residential home	*NA*	*NA*
- Nursing home	*NA*	*NA*
- Hospital ward	*NA*	*NA*
- Homeless	*NA*	*NA*
N (%) Cohabitees		
- Alone	*NA*	*NA*
- Spouse/partner	*NA*	*NA*
- Parent(s)	*NA*	*NA*
- Children under 18	*NA*	*NA*
- Children over 18	*NA*	*NA*
- Other family	*NA*	*NA*
- Non-family	*NA*	*NA*
N (%) victim of violence in the past year	*NA*	*NA*

Table B7: Primary analysis of the secondary outcomes

N; Mean (SD) [Range]	Intervention (N =)			Control (N =)			Partially Adjusted[1] Treatment Effect[3] (Intervention – Control) [95% CI] p-value	Fully Adjusted[2] Treatment Effect[3] (Intervention – Control) [95% CI] p-value
	Baseline	Month 10	Change	Baseline	Month 10	Change		
WEMWBS								
SWEMWBS								
Brief INSPIRE								
QPR-15								
TUS (hours in structured activity per week)								
ICECAP-A								
EQ-5D-5L								
Healthcare monitoring score[4]	NA		NA	NA		NA		

1 Adjusted for baseline and including random effects for GP practice

2 Adjusted for baseline as well as the stratification variables (practice size and locality) and including random effects for GP practice

3 Treatment effect is the between-group difference in the change in outcome between baseline and primary endpoint where available

4 Treatment effects presented as Incidence Rate Ratios

Table B8: Summary statistics of the secondary outcomes not subject to inferential analysis

N; Mean (SD) [Range]	Intervention (N =)			Control (N =)		
	Baseline	Month 10	Change	Baseline	Month 10	Change
N (%) Smoker			*NA*			*NA*
Number of times smoking per day						
N (%) Drinker			*NA*			*NA*
Number of times drinking per week						
N (%) Cannabis User			*NA*			*NA*
Healthcare Monitoring – Interaction with Primary Care						
# mental health consultations with GP	*NA*		*NA*	*NA*		*NA*
# mental health consultations with practice nurse	*NA*		*NA*	*NA*		*NA*
# physical health consultations with GP	*NA*		*NA*	*NA*		*NA*
# physical health consultations with practice nurse	*NA*		*NA*	*NA*		*NA*

# consultations with GP addressing physical and mental health	*NA*	*NA*	*NA*	*NA*
# consultations with practice nurse addressing physical and mental health	*NA*	*NA*	*NA*	*NA*
Healthcare Monitoring Outcomes - Surveillance				
N (%) Annual health check	*NA*	*NA*	*NA*	*NA*
N (%) Blood pressure recording	*NA*	*NA*	*NA*	*NA*
N (%) Weight recording	*NA*	*NA*	*NA*	*NA*
N (%) Lipids checked	*NA*	*NA*	*NA*	*NA*
N (%) Blood glucose checked	*NA*	*NA*	*NA*	*NA*
N (%) HbA1c checked	*NA*	*NA*	*NA*	*NA*
Healthcare Monitoring Outcomes – Active Management				
N (%) Changes/additions to blood pressure medication	*NA*	*NA*	*NA*	*NA*
N (%) Changes/additions to lipid medication	*NA*	*NA*	*NA*	*NA*

N %) Changes/additions to diabetes medication	*NA*	*NA*	*NA*	*NA*
N (%) Dietary advice given	*NA*	*NA*	*NA*	*NA*
N (%) Alcohol advice given	*NA*	*NA*	*NA*	*NA*
N (%) Smoking advice given	*NA*	*NA*	*NA*	*NA*
N (%) Smoking cessation intervention	*NA*	*NA*	*NA*	*NA*
Healthcare Monitoring Outcomes - Other				
N (%) Change to smoking status	*NA*	*NA*	*NA*	*NA*

Table B9: COVID-19 questionnaire

Outcome	Intervention (N=)	Control (N=)	Overall (N=)
N (%) in vulnerable group (received NHS letter to shield)			
- Yes - No - Don't Know			
N (%) extent of restrictions on usual activities			
- Not at all - Somewhat - Mixed - A lot - Extremely			
N (%) impact on mental health			
- Positively - Mixed - Negatively - Not at all			
N (%) impact on physical health			
- Positively - Mixed - Negatively - Not at all			
N (%) impact on relationships with others			
- Positively - Mixed - Negatively - Not at all			
N (%) impact on sleep routine			
- Positively - Mixed - Negatively - Not at all			
N (%) impact on finances			
- Positively - Mixed - Negatively - Not at all			

Table B10: COVID-19 sensitivity analysis 1 – excluding participants followed up on or after 23 March 2020

N; Mean (SD) [Range]	Intervention (N =)			Control (N =)			Partially Adjusted[1] Treatment Effect[3] (Intervention – Control) [95% CI] p-value	Adjusted[2] Treatment Effect[3] (Intervention – Control) [95% CI] p-value
	Baseline	Month 10	Change	Baseline	Month 10	Change		
MANSA								
TUS								

1 Adjusted for baseline and including random effects for GP practice

2 Adjusted for baseline as well as the stratification variables (practice size and locality) and including random effects for GP practice

3 Treatment effect is the between-group difference in the change in outcome between baseline and primary endpoint

Table B11: COVID-19 sensitivity analysis 2 – including an allocated group by pre/post COVID-19 lockdown interaction term

N; Mean (SD) [Range]	Intervention (N =)			Control (N =)			Treatment x COVID-19 interaction [95% CI] p-value
	Baseline	Month 10	Change	Baseline	Month 10	Change	
MANSA							
Pre-COVID[1]							
During COVID[2]							
Time Use Survey							
Pre-COVID[1]							
During COVID[2]							

1: Participants who provided primary outcome follow-up data before 23 March 2020

2: Participants who provided primary outcome follow-up data on or after 23 March 2020

Table B12: COVID-19 sensitivity analysis 3 – adjusting for pre/post COVID-19 lockdown indicator

N; Mean (SD) [Range]	Intervention (N =)			Control (N =)			COVID-19 Coefficient [95% CI] p-value	Fully Adjusted[1] Treatment Effect (Intervention – Control) [95% CI] p-value
	Baseline	Month 10	Change	Baseline	Month 10	Change		
MANSA								
Pre-COVID[2]								
During COVID[3]								
Time Use Survey								
Pre-COVID[2]								
During COVID[3]								

1: Adjusted for baseline MANSA as well as the stratification variables (practice size and locality) and including random effects for GP practice
2: Participants who provided primary outcome follow-up data before 23 March 2020
3: Participants who provided primary outcome follow-up data on or after 23 March 2020

Table B13: Safety Outcomes

	Intervention Participants with at least one care partner interaction	**Participants without any care partner interaction**
Number of psychiatric hospital admissions		
Number of days inpatient as a result of psychiatric admission		
Number of crisis care (home treatment) episodes		
Number of days under home treatment (crisis care)		
Mean number of psychiatric hospital admissions per participant		
Mean days inpatient as a result of psychiatric admission per participant		
Mean number of crisis care (home treatment) episodes per participant		
Mean number of days under home treatment (crisis care) per participant		

Table B14: EQ-5D responses, participants with complete cost and QALY data

EQ-5D health states	**Intervention n/N (%; 95% CI)**		**Comparator n/N (%; 95% CI)**	
Mobility	Baseline	10-month	Baseline	10-month
No problems				
Slight problems				
Moderate problems				
Severe problems				
Extreme problems				
Self care				
No problems				
Slight problems				
Moderate problems				
Severe problems				
Extreme problems				
Usual activities				
No problems				
Slight problems				
Moderate problems				
Severe problems				
Extreme problems				
Pain and distress				
No problems				
Slight problems				
Moderate problems				
Severe problems				
Extreme problems				
Anxiety and depression				
No problems				
Slight problems				
Moderate problems				
Severe problems				
Extreme problems				

Table B15: ICECAP-A responses, participants with complete cost and QALY data

EQ-5D health states	Intervention n/N (%; 95% CI)		Comparator n/N (%; 95% CI)	
	Baseline	10-month	Baseline	10-month
1 Feeling settled and secure				
I am able to feel settled and secure in all areas of my life				
I am able to feel settled and secure in many areas of my life				
I am able to feel settled and secure in a few areas of my life				
I am unable to feel settled and secure in any areas of my life				
2 Love and Friendship				
A lot of love, friendship and support				
Quite a lot of love, friendship and support				
A little love, friendship and support				
No love, friendship and support				
3 Independence				
I am able to be completely independent				
I am able to be independent in many things				
I am able to be independent in a few things				
I am unable to be at all independent				
4 Achievement and progress				
I can achieve and progress in all aspects of my life				
I can achieve and progress in many aspects of my life				
I can achieve and progress in a few aspects of my life				
I cannot achieve and progress in any aspects of my life				
5 Enjoyment and pleasure				
I can have a lot of the enjoyment and pleasure				
I can have quite a lot of enjoyment and pleasure				
I can have a little enjoyment and pleasure				
I cannot have any enjoyment and pleasure				

Table B16: Average utility values and QALYs, baseline to 10-month follow up

Measure	Complete case		Imputed	
	Intervention	Comparator	Intervention	Comparator
	Mean, sd (95% CI)	Mean, sd (95% CI)	Mean, sd (95% CI)	Mean, sd (95% CI)
EQ-5D-5L				
Utility				
QALY				
ICECAP-A				
Utility				
QALY				

Table B17: Average costs for 3-month participant reported data, complete case, primary analysis

Cost category	Costs for 3-month recall		Costs for extrapolated 10-month period	
	Intervention	Comparator	Intervention	Comparator
	Mean, sd (95% CI)	Mean, sd (95% CI)	Mean, sd (95% CI)	Mean, sd (95% CI)
Primary care (excluding GP and practice nurse)				
Community care				
Social care				
Non-psychiatric inpatient care				
Hospital outpatient and day visits				
Emergency department visits				

Table B18: Average costs by cost category, baseline to 10-month follow up

Cost category	Complete case		Missing follow-up and observations imputed	
	Intervention	Comparator	Intervention	Comparator
	Mean, sd (95% CI)	Mean, sd (95% CI)	Mean, sd (95% CI)	Mean, sd (95% CI)
Primary care				
Community care				
Social care				
Home-based crisis care				
Hospital inpatient stays				
Hospital outpatient and day visits				
Emergency department visits				

Table B19: Net costs and QALYs, imputed data, baseline to 10-month follow up

Analysis	Net cost	Net QALY	ICER	Probability intervention is cost effective
	Mean, se, 95% CI	Mean, se, 95% CI	Point estimate	Point estimate
Primary analysis				
Sensitivity analyses				
Alternative measures of health benefit				
MANSA				
ICECAP-A				
Estimating 10-month costs for 3-month participant report service use				
Constant monthly cost				
Increasing monthly cost for 3 months post baseline				
Accounting for missing data				
Complete case analysis				
Indicator analysis				
Single imputation				

APPENDIX C: CONSORT flow diagram of GP practices/participants