Reviewing long-term antidepressant use by careful monitoring in everyday practice (REDUCE)

Statistical Analysis Plan

Trial registration number:	https://www.isrctn.com/ISRCTN12417565 Registered 07/10/2019
Protocol title and version number:	https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-020-04338-7
SAP version number:	1
SAP date:	21 April 2022

To be approved and reviewed by:

	Name	Signature	Date
Chief Investigator	Professor Tony Kendrick		21/4/2022
Trial Statistician (author)	Dr Beth Stuart		21/4/2022

Table of Contents

1 Introduction

1.1 Purpose of SAP

This statistical analysis plan (SAP) describes in detail the methods that will be used to analyse the data collected as part of the REDUCE trial. This will form the basis of the final trial publication. The final analysis will follow the SAP to ensure that the analyses are conducted in a scientifically valid manner and to avoid post hoc decisions which may affect the interpretation of the statistical analysis. Any deviations from the SAP will be detailed in the final report.

1.2 Trial background and rationale (from ISRCTN registration)

Work Stream 5 (WS5) of the REDUCE programme aims to determine the effectiveness of online (Internet) interventions which support practitioners and guide patients on coming off antidepressants, together with psychological practitioner telephone calls to support the patients.

We will assess the effectiveness of the interventions in terms of reductions in antidepressant use in the absence of worsening of depression, and assess patients' and practitioners' use of the interventions (automatically recorded by the Southampton 'LifeGuide' software used for the Internet guidance).

1.3 Objectives (from published protocol[1])

1. To determine the effectiveness of the intervention through a randomised controlled trial
2. To estimate cost-effectiveness from a health and personal social service perspective, with a sensitivity analysis from a societal perspective.

   The analysis required to evaluate the first of these objectives will be set out in this Statistical Analysis Plan. The analysis required to evaluate the second objective will be set out in a separate Health Economics Analysis Plan.

1.4 Definition of endpoints

1.4.1 Definition of primary endpoint

The PHQ-9 scores for depressive symptoms at 6 months.

1.4.2 Definition of secondary endpoints

- The PHQ-9 score over the full 12 month period as a repeated measure analysis
- Discontinuation of antidepressants at 6 months
- Antidepressant withdrawal symptoms measured on the Discontinuation Emergent Signs and Symptoms scale (DESS) over the trial period
- Anxiety on the GAD-7 measure of generalised anxiety disorder over the trial period
- Wellbeing scores over the trial period on the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)
- Modified version of the Medical Informant Satisfaction Scale (MISS) over the trial period to measure patient satisfaction over the follow-up period.
- Patient enablement measured by the Patient Enablement Instrument (PEI) over the trial period

1.5 Analysis principles

All analyses will be reported in line with thee CONSORT 2010 extension to cluster randomised trials[2] and non inferiority trials[3] and Southampton Clinical Trials Unit (SCTU) standard operating procedure (SOP) on planning, implementing and reporting statistical analyses (CTU/SOP/5058).

2 Design considerations

2.1 Description of trial design

This is a pragmatic multicentre cluster randomised non-inferiority trial.

2.2 Trial power and sample size

To have 90% power, with a one-sided alpha of 2.5%, to establish non-inferiority in terms of depressive symptoms within two points (estimated to be the minimal clinically important difference) on the PHQ-9 at 6 months (standard deviation 5.4), we need 155 patients followed up in each arm. Assuming a variable cluster size of between 1 and 7 per practice (mean 3) and an intra-cluster correlation coefficient of 0.012 (from the Health Technology Assessment THREAD trial of treating mild-to-moderate depression in primary care[4]), gives a 1.033 cluster design effect (based on a coefficient of variation of 1.5/3 = 0.5, using the formula of Eldridge[5]). Anticipating 20% do not comply with the intervention and/or are lost to follow-up, we need to randomise (155 × 2 × 1.036) / 0.8 = 402 patients (201 per arm) from 134 practices (67 per arm).

Following discussion with the study steering committees and the funder, the sample size was amended to 360 participants. This allows for a correlation between baseline and follow-up scores with a deflation factor of $1-p^2$. At the time of review we observed a correlation of 0.47 but were unsure whether this would persist to the end of follow up for all participants. If we assume the more conservative estimate of correlation of 0.26 (the bottom end of the CI), the target sample size to achieve 90% power would 375. We assume the final figure will be somewhere in between 0.26 and 0.47, and that a sample of around 360 will therefore provide us with the necessary 90% power to test reliably for non-inferiority of the intervention in terms of depressive symptoms.

2.3 Randomisation details

Randomisation of practices is by computerised sequence generation, and minimisation with a random element using three factors to avoid imbalance between the two arms: practice size (large/small), location (urban/rural) and social deprivation (dichotomised around the median Index of Multiple Deprivation score). The allocation ratio is 1:1 but there is a random element to the minimisation algorithm and so we might not expect perfect balance to the randomisation.

2.4 Timing of planned analyses

2.4.1 Interim analyses and early stopping

No interim analysis is planned and no pre-specified stopping rules have been established.

2.4.2 Final analysis

End of study is defined as when the last patient has had their last data collected, cleaned and verified.

3 Statistical considerations

3.1 Definition of analysis populations

3.1.1 Intention-to-treat analysis population

This population includes all randomised practices and all patients recruited within them regardless of treatment compliance. This includes the participants recruited during WP4 and treated as an internal pilot. All summaries and analysis will be on the modified ITT population unless otherwise specified, i.e. the population as randomised but without missing data imputed. The ITT analysis will be the primary analysis.

3.1.2 Per-protocol analysis population

In a non-inferiority trial where some patients do not comply with treatment as randomised, the difference between the arms can appear reduced and the groups look more similar, leading to the incorrect conclusion of non-inferiority. A per-protocol analysis would analyse individuals based on their compliance with treatment as randomised, excluding non-compliant participants, which may give a more conservative estimate of effect for non-inferiority. However, the exclusion of some participants after randomisation can potentially lead to bias. Therefore, we will present both intention-to-treat and per-protocol analyses. The ITT analysis will be the primary and the per-protocol analyses will be treated as secondary.

3.1.3 CACE analysis population

We will also undertake a complier-average causal effect analysis, which is another approach for dealing with non-compliance that compares compliant participants in the intervention group with those in the control group whose characteristics are similar enough to the intervention group compliers to suggest they too would have complied with the intervention given the opportunity to do so.[6] Compliance for these analyses in the intervention arm will be defined as completing the first session of the LifeGuide programme within 6 months of recruitment (anticipating >90%). The first session will have information about antidepressant treatment, the rationale for attempting withdrawal, and how withdrawal should be attempted under supervision. We would expect patients to benefit from that session even if they do not log on again.

3.2 Analysis software

SAS v9.4 or higher, or Stata v15.1 or higher will be used for all analyses.

3.3 Methods for handling data

3.3.1 Withdrawal from trial

All data up until the point of patient withdrawal from the trial will be used in analyses unless the patient withdrew consent and does not wish for the data already collected prior to withdrawal to be used for the trial.

If a practice withdraws from the trial, no further patients will be recruited. All data on patients collected until that point will be used and any patients recruited will continue to be followed up in accordance with the trial schedule.

3.3.2 Missing data

The primary analysis will be of complete cases.

If more than 2 items in the PHQ-9 and the GAD7 have missing values, the total score will be missing. If one or two items are missing, the score will be imputed with the mean of the non-missing scores before summing.

We will examine the structure and pattern of missing data and, if appropriate, will present a sensitivity analysis based on data imputed using a chained equations multiple imputation model. The imputation model would include the outcome measure, baseline value of the outcome, randomisation group, clustering by practice and all covariates included in the analysis model (see below)

3.3.3 Outliers

No methods will be used to handle outliers in the data, except in the regression models. If outliers are found then firstly the source data will be checked. If the source data is correct, then a sensitivity analysis will be performed excluding them from the analysis.

3.3.4 Assumption checking and alternative methods

Assumptions for linear regression models (linearity, normality, homoscedasticity) will be checked using scatter plots of standardized residuals against fitted values, and qq plots. If linear models are not appropriate a log-linear transformation will be used.

3.4 Definition of key derived variables

The PHQ-9[7] is a self-complete questionnaire taking approximately 3 min to complete. Each item is scored 0-3 with a total possible score ranging from 0-27. Higher scores indicate more severe symptoms.

The Generalised Anxiety Disorder Assessment 7-item version[8] is a self-report measure of anxiety symptoms. Each item is scored from 0-3 with a total possible score ranging from 0-21. Higher scores indicate more severe symptoms.

Discontinuation of antidepressants is deemed to have occurred once the patient has stopped taking them for 2 months. This is a binary outcome (did not discontinue/did discontinue).

The Warwick–Edinburgh Mental Wellbeing Scale (WEMWBS)[9] measures both subjective experiences of happiness and life satisfaction (the 'hedonic perspective'), and positive psychological functioning, good relationships with others and self-realisation (the 'eudaimonic perspective'). Each item is answered on a 1 to 5 Likert scale and summed to give a total score ranging form 14 to 70. Higher scores indicate greater wellbeing.

Antidepressant withdrawal symptoms are measured using the Discontinuation Emergent Signs and Symptoms scale[10], a brief self-report measure on which participants can indicate the presence of, and

changes in, 43 possible antidepressant withdrawal symptoms. The total number of symptoms is summed.

The 29-item 'Medical Interview Satisfaction Scale' (MISS-29) was developed in the USA to assess patient satisfaction with individual doctor-patient consultations and has been shown to be valid and reliable in UK primary care[11].

The modified Patient Enablement Instrument[12] asks 6 questions about the patients' ability to understand their problems and cope with illness. These are rated on a 7 point Likert scale (0=strongly agree to 7=strongly disagree). A mean overall score can be calculated, with lower scores equating to better enablement.

3.5 General principles for reporting and analysis

The following general principles for reporting and analysis will be used:

- 5% two-sided level of statistical significance, with corresponding 95% confidence intervals presented where applicable.
- No adjustments for multiplicity are planned.
- Summary statistics will include either mean, standard deviation, and/or median, interquartile range.
- Treatment groups will be labelled in the tables as Intervention Group and Control Group accordingly, and a total column will be included in tables were applicable.

4 Planned analyses and reporting

4.1 Disposition of the study population

A CONSORT diagram (see Appendix 1) will be produced showing a clear account of all practices and patients who entered the trial- see below for an example figure. Withdrawal information including the primary reasons of discontinuation will be summarised and presented by where this is known.

4.2 Protocol deviations

A listing of all Major or Potential/Serious Breach (Major protocol deviations with potential to affect patient safety/data) and Potential/Serious Breach (with actual affect to patient safety/data, Major/Potential/Serious Breach of GCP guidelines or consistent non-compliance by site) will be produced (by patient and site where applicable)

4.3 Baseline and demographic characteristics

Summary statistics will be produced and presented by group for demographic and baseline characteristics but no comparisons will be undertaken, rather the clinical importance of any imbalance will be noted. If there are imbalances of clinical importance in variables not listed in 4.4 below, we will control for these as covariates in the analyses.

4.4 Primary endpoint

The primary outcome, that is, the differences at 6 months between intervention and control in depression as measured by the PHQ-9, will be analysed using a linear mixed model, adjusting for duration of depression, past history of depression, age, gender, marital status, no. of dependents, ethnic group, education level, economic position (employment status), and accommodation status, baseline depression score, anxiety score, internet use and clustering, including practice as a random effect. These covariates have been chosen based on their known relationship with the outcome from previous literature.

The model will use all the observed data and makes the assumption that missing PHQ-9 scores are missing completely at random. We will examine the lower limit of the 95% confidence interval to establish non-inferiority.

We will then undertake the same analyses using the per protocol population and the CACE population.

Whilst the ITT analysis will be the primary analysis, we will report the analyses based on all three approaches and interpret the findings cautiously in light of any differences between approaches that may emerge. This analysis will form the core of the publication of the trial results.

4.5 Secondary endpoints

The discontinuation of antidepressants will be evaluated at 6 months as a binary outcome in a mixed logistic regression model, using the same modelling approach and covariates as for the primary outcome.

Analysis of all other secondary outcomes will also be conducted using generalised linear mixed regression models adjusting for stratification variables, the potential confounders listed in 4.4 above and the baseline value of the outcome. These will be evaluated as repeated measures so all models will allow for the clustering of observations within participants over time and of participants within practices.

4.6 Additional analyses

No subgroup analyses are planned. Any post-hoc analyses will be exploratory only. Health economic analyses will be undertaken and a separate Health Economics Analysis Plan will be prepared.

In accordance with the CONSORT recommendations for cluster randomised trials, we will also report the ICC for the primary outcome.

Due to the COVID-19 pandemic and subsequent lockdown period, it is possible that there may be changes to the key outcomes unrelated to randomisation group. We will therefore look at the scores in each arm in the pre-, peri- and post-COVID periods in the whole study population. We will use descriptive statistics and graphical representations to explore any trends and aim to control for any time varying effect on outcomes in a sensitivity analysis

4.7 Safety reporting

A listing of serious adverse events (SAEs) will be provided for all related/unrelated SAEs. If required, a summary table will also be presented.

5 References

1. Kendrick, T. *et al.* REDUCE (Reviewing long-term antidepressant use by careful monitoring in everyday practice) internet and telephone support to people coming off long-term antidepressants: Protocol for a randomised controlled trial. *Trials* **21**, 1–15 (2020).

2. Campbell, M. K., Piaggio, G., Elbourne, D. R. & Altman, D. G. Consort 2010 statement: Extension to cluster randomised trials. *BMJ* **345**, (2012).

3. Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J. W. & Altman, D. G. Reporting of Noninferiority and Equivalence Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA* **308**, 2594–2604 (2012).

4. Kendrick, T. *et al.* Randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of selective serotonin reuptake inhibitors plus supportive care, versus supportive care alone, for mild to moderate depression with somatic symptoms in primary care: the THREAD (THREshold for AntiDepressant response) study. *Health Technol. Assess.* **13**, iii–iv, ix (2009).

5. Eldridge, S. M., Ashby, D. & Kerry, S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int. J. Epidemiol.* **35**, 1292–1300 (2006).

6. Angrist, J. D., Imbens, G. W. & Rubin, D. B. Identification of Causal Effects Using Instrumental Variables. *J. Am. Stat. Assoc.* **91**, 444–455 (1996).

7. Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O. & Löwe, B. Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Ann. Intern. Med.* **146**, 317–325 (2007).

8. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch. Intern. Med.* **166**, 1092–1097 (2006).

9. Stewart-Brown, S. *et al.* The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. *J Epidemiol Community Heal.* **65**, A38–A39 (2011).

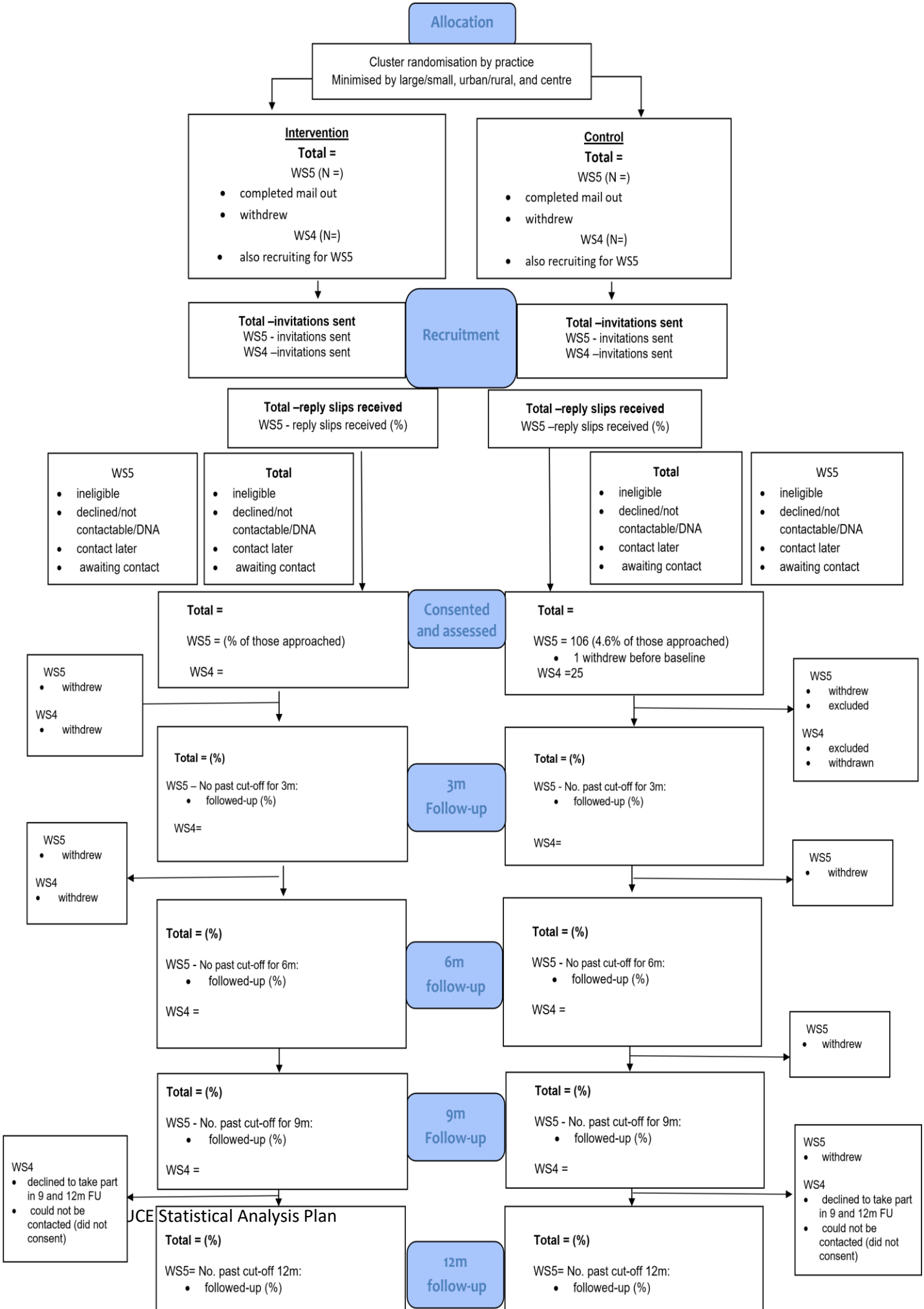
10. Rosenbaum, J. F., Fava, M., Hoog, S. L., Ascroft, R. C. & Krebs, W. B. Selective serotonin reuptake inhibitor discontinuation syndrome: a randomized clinical trial. *Biol. Psychiatry* **44**, 77–87 (1998).

11. Meakin, R. & Weinman, J. The 'Medical Interview Satisfaction Scale' (MISS-21) adapted for British general practice. *Fam. Pract.* **19**, 257–263 (2002).

12. Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain. doi:10.1136/bmj.a884.

Appendix 1

REDUCE RCT Consort diagram


Allocation
Cluster randomisation by practice
Minimised by large/small, urban/rural, and centre
Intervention
Total =
WS5 (N =)
completed mail out
withdrew
WS4 (N=)
also recruiting for WS5
Control
Total =
WS5 (N =)
completed mail out
withdrew
WS4 (N=)
also recruiting for WS5
Total –invitations sent
WS5 - invitations sent
WS4 –invitations sent
Recruitment
Total –invitations sent
WS5 - invitations sent
WS4 –invitations sent
Total –reply slips received
WS5 - reply slips received (%)
Total –reply slips received
WS5 –reply slips received (%)
WS5
ineligible
declined/not contactable/DNA
contact later
awaiting contact
Total
ineligible
declined/not contactable/DNA
contact later
awaiting contact
Total
ineligible
declined/not contactable/DNA
contact later
awaiting contact
WS5
ineligible
declined/not contactable/DNA
contact later
awaiting contact
Total =
WS5 = (% of those approached)
WS4 =
Consented and assessed
Total =
WS5 = 106 (4.6% of those approached)
1 withdrew before baseline
WS4 =25
WS5
withdrew
WS4
withdrew
WS5
withdrew
excluded
WS4
excluded
withdrawn
Total = (%)
WS5 – No past cut-off for 3m:
followed-up (%)
WS4=
3m Follow-up
Total = (%)
WS5 - No. past cut-off for 3m:
followed-up (%)
WS4=
WS5
withdrew
WS4
withdrew
WS5
withdrew
Total = (%)
WS5 - No past cut-off for 6m:
followed-up (%)
WS4 =
6m follow-up
Total = (%)
WS5 - No past cut-off for 6m:
followed-up (%)
WS4 =
WS5
withdrew
Total = (%)
WS5 - No. past cut-off for 9m:
followed-up (%)
WS4 =
9m Follow-up
Total = (%)
WS5 - No. past cut-off for 9m:
followed-up (%)
WS4 =
WS4
declined to take part in 9 and 12m FU
could not be contacted (did not consent)
WS5
withdrew
WS4
declined to take part in 9 and 12m FU
could not be contacted (did not consent)
Total = (%)
WS5= No. past cut-off 12m:
followed-up (%)
12m follow-up
Total = (%)
WS5= No. past cut-off 12m:
followed-up (%)