Bristol Randomised Trials Collaboration (BRTC)

WISE

Wellbeing in Secondary Education

Statistical Analysis Plan

Version 2.5 (25/09/2018)

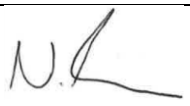
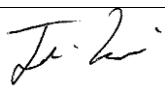
The following people have reviewed the Statistical Analysis Plan and are in agreement with the contents			
Name	**Role**	**Signature**	**Date**
Bryar Kadir and Nicholas Turner	Author		25/09/2018
Karla Hemming	Statistical Reviewer		27 /9/2018
Judi Kidger	Chief Investigator		25/09/2018

Table of Contents

1. INTRODUCTION & PURPOSE 7
2. SYNOPSIS OF STUDY DESIGN AND PROCEDURES 7
2.1. Trial objectives and aims 7
2.1.1. Primary objective 7
2.1.2. Secondary objectives 7
2.2. Trial design and configuration 8
2.3. Trial centres 8
2.4. Eligibility criteria 8
2.4.1. Cluster inclusion criteria 8
2.4.2. Cluster exclusion criteria 8
2.4.3. Individual inclusion criteria 8
2.4.4. Individual exclusion criteria 9
2.5. Description of interventions 9
2.6. School Selection and Randomisation procedures 9
2.7. Sample size and justification 10
2.8. Blinding and breaking of blind 10
2.9. Trial committees 10
2.10. Outcome measures 11
2.10.1. Primary outcome 11
2.10.2. Secondary outcomes 11
3. GENERAL ANALYSIS CONSIDERATIONS 12
3.1. Analysis datasets 12
3.2. Procedures for missing data 12
3.3. Cluster effects 13
3.4. Outliers 13
3.5. Visit windows 13
3.6. Definition 13
4. DESCRIPTION OF PARTICIPANT CHARACTERISTICS 13
4.1. Disposition 13
4.2. Baseline characteristics 14
5. ASSESSMENT OF STUDY QUALITY 14
5.1. Eligibility checks 14

5.2. Study completion 14
5.3. Protocol deviations 14
5.4. Specify & justify changes made to the planned statistical analyses 14
6. ANALYSIS OF EFFECTIVENESS 14
6.1. Summary of primary and secondary outcomes 14
6.2. Primary analysis 15
6.3. Secondary analyses 15
6.4. Sensitivity analysis 17
6.5. Subgroup analysis 17
6.6. Exploratory analysis 18
7. ANALYSIS OF SAFETY 20
8. FINAL REPORT TABLES AND FIGURES 21
8.1. Subject characteristics and background summaries 21
8.2. Study quality summaries 23
8.3. Outcome summaries 23
8.4. Primary outcome results 24
8.5. Secondary outcomes results 25
8.6. Sensitivity analysis for primary endpoint 26
8.7. Sensitivity analysis for secondary endpoints 26
8.8. Subgroup analysis 27
8.9. Safety results 28
9. APPENDICES 28
9.1. Stata code for derived variables 28
9.2. Stata code for final analyses 28
9.3. Details of standard assessment tools 28
10. REFERENCES 28

List of tables

Table 1. Baseline Demographics - Teachers…… 17

Table 2. Baseline Demographics - Students…… 18

Table 3. Baseline Demographics – Schools…..……18

Table 4. Consort Diagram…… 19

Table 5. Primary endpoint summary…….. 20

Table 6. Secondary endpoint summary…….. 20

Table 7. Primary outcome result…….. 20

Table 8. Secondary outcome results..…….. 21

Table 9. Sensitivity analysis - primary endpoint results.…….. 22

Table 10. Sensitivity analysis - secondary endpoint results……...22

Table 11. Subgroup analysis - Subgroup analysis results…….. 23

Table 12. Adverse events…….. 24

Abbreviations

ALPHA	A young people's research advisory group
BRTC	Bristol Randomised Trial Collaboration
CACE	Complier Average Causal Effect
CAMHS	Child and Adolescents Mental Health Services
DECIPHer	Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement
FG	Focus group
FREC	Faculty Research Ethics Committee
FHSEC	Faculty of Health Sciences Ethics Committee
FSM	Free School Meal status
HSC	Healthy Schools Coordinator
ICC	Intraclass Correlation
ID	Identity
INSET	IN-SErvice Training
LA	Local Authority
MICE	Multivariate Imputation by Chained Equations
MHFA	Mental Health First Aid
MAR	Missing At Random
MRC	Medical Research Council
Ofsted	Office for Standards in Education
OECD	Organisation for Economic Co-Operation and Development
PHQ-8	Patient Health Questionnaire 8 item
PI	Principle Investigator
PMG	Project Management Group
PS	Peer supporter
Participant	An individual who takes part in a trial
NIHR	National Institute for Health Research
NIHR PHR	National Institute for Health Research Public

	Health Research
RCT	Randomised Controlled Trial
SAE	Serious Adverse Event
SDQ	Strengths and Difficulties Questionnaire
SHRN	School Health Research Network
SLT	Senior leadership teams in schools
SL	Senior leaders in schools
STARS	Supporting Teachers and Children in Schools
T0	Baseline data collection period
T1	Follow up data collection period, 12 months after baseline
T2	Follow up data collection period, 24 months after baseline
TSC	Trial Steering Committee
WISE	Wellbeing in Secondary Education
WEMWBS	Warwick Edinburgh Mental Wellbeing Scale
WPAI	Work Productivity and Activity Impairment Questionnaire
YMHFA	Youth Mental Health First Aid

1. INTRODUCTION & PURPOSE

This document details the rules proposed and the presentation that will be followed, as closely as possible, when analysing and reporting the main results from the WISE study: A cluster randomised trial of a mental health training and support intervention provided to secondary school teachers.
The purpose of the plan is to:

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

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

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

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

2. SYNOPSIS OF STUDY DESIGN AND PROCEDURES

The information in these sections has been extracted from the protocol on 18 May 2016 in order to place the analysis plan within the context of the trial aims and methods.

2.1. Trial objectives and aims

2.1.1. Primary objective

To establish if the WISE intervention leads to improved teacher emotional wellbeing compared to usual practice.

2.1.2. Secondary objectives

1. Does the WISE intervention lead to lower levels of teacher depression, absence and presenteeism, improved student wellbeing, attendance and attainment, and reduced student mental health difficulties compared to usual practice?

2. Do any effects of the intervention differ according to the proportion of children receiving free school meals (FSM – an indicator of the socioeconomic catchment area) and geographical area, or individual level baseline mental health, gender, ethnicity and FSM?
3. What is the cost of the WISE intervention, and is it justified by improvements to staff and student wellbeing and reductions to staff depression and student difficulties?
4. Does the WISE intervention work according to the mechanisms of change hypothesised in the logic model?
5. Is the WISE intervention sustainable?

2.2. Trial design and configuration

This is a cluster randomised controlled trial, with schools as the unit of allocation.

2.3. Trial centres

The study will be conducted within two geographical areas (Bristol/ surrounding areas of Bristol and South Central/ South East Wales).

2.4. Eligibility criteria

2.4.1. Cluster inclusion criteria

All state mainstream secondary schools within the relevant geographical areas.

2.4.2. Cluster exclusion criteria

I. Fee paying schools
II. Special schools (e.g. for those with learning disabilities)
III. Pupil referral units
IV. Schools that were pilot schools
V. Schools already participating in other, similarly intensive, research studies (healthy schools teams will advise on this)
VI. Schools that have already provided or are providing MHFA training to teachers
VII. Schools that do not have FSM data
VIII. Schools in the same academy chain and LA as another participating school

2.4.3. Individual inclusion criteria

All staff who are in teaching posts.

2.4.4. Individual exclusion criteria

All staff who are in non-teaching posts.

2.5. Description of interventions

Each intervention school will receive the following:

- Staff peer support service: 8 % of staff who are nominated via confidential voting of questionnaire respondents will be selected (ensuring a mixture of gender and years of experience and (non)teaching role) and will receive 2 day MHFA training. Then will act as peer supporters to other staff members.
- INSET teacher training using MHFA for schools and colleges: Another 8 % of teachers only (targeted at tutors or heads of year) who did not receive the 2 day MHFA training, will receive one day youth MHFA course. The learning will be applied during usual interactions with students.
- Mental health awareness raising session: One hour training for all teachers, during in-service training or meeting time, that will introduce the intervention and focus on the importance of mental health issues in school.

Control schools will continue with usual practice.

2.6. School Selection and Randomisation procedures

Secondary schools within each country will be organised into four areas (Bristol / surrounds of Bristol in England and South East / South Central in Wales). Within each area grouping, the schools will be stratified into three levels according to FSM (high, medium and low). In Wales, two schools will be randomly selected from each stratum in each area; this will be conducted by statisticians blinded to any identifying information. In England, all schools will be invited for expressions of interest to participate, and where more than two in each stratum respond, two will be randomly selected. If it is not possible to obtain consent to participate from 2 schools within one FSM stratum then it will be joined to another stratum and 4 schools will be picked from that stratum. For example if 2 high FSM schools cannot be recruited, then high and medium FSM strata will be joined and 4 schools selected from this new stratum.

The schools that have been selected to participate in the trial (as described above) will be randomly allocated to either the control or intervention arm of the trial stratified by geographical region (Bristol area of England, and South East/South Central Wales) and FSM level (high, medium or low).

2.7. Sample size and justification

The primary outcome measure will be the WEMWBS scale. A change of 3 points on this scale is close to the difference in mean scores between the highest and lowest ranked schools in the pilot study. The sample size takes account of clustering; in the pilot data from 6 schools, the WEMWBS ICC (calculated using the loneway command in Stata) for teachers was 0.01 (95%CI 0.00, 0.05) [1]. We assume a mean of 50 teachers followed up per school after allowance for 20% dropout (with a coefficient of variation of sample size of 0.5), and an SD for WEMWBS of 8.4 (based on the pilot data). Given that estimates based on pilot data can be unreliable [2], assuming an ICC of 0.05, which is greater than that seen in the pilot, -a sample size of 24 schools (12 intervention and 12 control) would achieve over 80% power. This would rise to 98% if the ICC is 0.02, and fall to 68% for an ICC of 0.08. These estimates of power are likely to be conservative given that the sample size calculation does not account for the baseline assessment of the outcome.

2.8. Blinding and breaking of blind

Allocation to study arm will take place after baseline data are collected to ensure blinding among all parties during this first data collection. It is not possible for the schools, teachers or students to be blind to intervention status. The research assistants/associate leading the outcome data collection, will also be collecting the process data which will prevent blinding. The schools will however be assigned ID numbers and as such casual staff and the statisticians analysing the primary and secondary outcomes data will be blind to the schools' identities and intervention status. A timeline diagram of blinding status throughout the trial processes will be produced for clarification [3].

2.9. Trial committees

The study team will hold 2-4 monthly Project Management Group (PMG) meetings, along with the researchers, statisticians and health economists working on the project, and two members of the public (teachers) to plan each phase of the study.

The Trial Steering Committee (TSC) will be made up of an independent chair with statistical and trial expertise, the study PI, a co-applicant from the Cardiff site, an independent statistician, two further independent scientists with relevant expertise and the Head of Research and Policy at the Education Support Partnership. Meetings will also be attended by other co-applicants or study team members as necessary. The TSC will monitor trial progress and conduct, and advise on scientific credibility. It will meet annually throughout the study, with teleconference and email contact in the interim as necessary.

2.10. Outcome measures

2.10.1. Primary outcome

The primary outcome is the average effect of the intervention on mental wellbeing of teachers over the duration of follow up, measured using the self-completed WEMWBS questionnaire [4] collected at baseline, one and two year follow ups. The WEMWBS scales from 14-70 and will be analysed as a continuous variable.

2.10.2. Secondary outcomes

Time averaged treatment effects will be reported for those variables collected at baseline, one and two year follow ups:

- Teacher depression measured by PHQ-8 [5] that scales from 0 to 24
- Teacher presenteeism measured by WPAI [6], that scales from 0 to 10
- Teacher absence (ever/never in the past 4 weeks) and total number of days of absence, which will be collected at baseline, one and two years follow up by self-report

Adjusted difference between trial arms at 2 years follow up will be reported for those variables collected only at baseline and two year follow up:

- Student mental wellbeing measured by the WEMWBS
- Student psychological distress, measured by the SDQ [7] that scales from 0 to 40
- Teacher absence collected from schools' routine data sources
- Teacher retirement/resignation collected from schools' routine data sources
- Student attendance (%) for all year groups
  Year 11 attainment (percentage achieving five or more grades 5-9 GSCEs including English and Maths)

Teacher PHQ-8 and student WEMWBS and SDQ will be measured as continuous variables. PHQ-8 will also be measured as an ordinal variable (a score of 0-4 indicating no depressive symptoms / 5-10 indicating mild symptoms / 10-14 indicating moderate symptoms / 15-19 indicating moderately severe symptoms 20-24 indicating severe) and a binary variable, with a cut-point of 10 or more indicating depression. Teacher absence and presenteeism will be treated as binary (any vs no absence in the previous four working weeks and health problems having 0 effect vs 1-10 effect on work over the previous four working weeks) to ascertain the impact of the intervention on any absence or presenteeism. Teacher absence will also be categorised as 0 / 1 day or less / 2 -7 days / more than 7 days, and the categories for teacher presenteeism will be 0 / 1-5 / 6-10. Teacher absence and teacher retirements/resignations as reported by schools, student attendance and

student attainment (% students achieving five or more 5-9 grade GCSEs including English and Maths) will all be reported at a school level and will be treated as continuous variables.

	Baseline	Year 1	Year 2
Teacher	• WEMWBS • PHQ-8 • Presenteeism • Teacher absence (past 28 days)	• WEMWBS • PHQ-8 • Presenteeism • Teacher absence (past 28 days)	• WEMWBS • PHQ-8 • Presenteeism • Teacher absence (past 28 days)
Student	- WEMWBS - SDQ		- WEMWBS - SDQ
School – Teacher	- Teacher absence (%) - Retirement/resignation		- Teacher absence (%) - Retirement/resignation
School – Student	- Attendance (%) - Attainment (%)		- Attendance (%) - Attainment (%)

3. GENERAL ANALYSIS CONSIDERATIONS

3.1. Analysis datasets

Full Analysis set: All teachers who are in post and all students in the appropriate year group who are on the register, in the randomised schools, during at least one of baseline, one or two year data collection, will form the main analysis sample. It is analysis of this sample which is in accordance with the "intention to treat" (ITT) principle.

Per protocol set: All participants in the Full Analysis set who are from schools with no major protocol violations that could interfere with the objectives of the study.

ITT will be used as primary outcome, then as sensitivity we will use ITT and impute missing data, and we will also present per protocol analysis, as well as CACE analysis.

3.2. Procedures for missing data

Missing data for teacher WEMWBS and PHQ-8 outcomes and student WEMWBS and SDQ outcomes (at baseline or at any follow-up) will be imputed using one or more appropriate imputation techniques depending on the patterns of missing data. If multiple imputation is used, imputation

models will include all outcome timepoints (baseline and every follow-up), intervention arm, stratifying variables and a random effect for school to allow for clustering, as well as any appropriate baseline/ auxiliary covariates that are indicative of missingness.

3.3. Cluster effects

The analyses will take account of clustering by school and clustering due to repeated measures using random effects. If models fail to converge then robust standard errors (which account for intragroup correlation) will be used instead of random effects.

3.4. Outliers

The data will be checked for validity, each variable will be examined separately, and any outliers (>3SD of the mean) will be checked for data entry errors. Where no error is identified, the variable will be checked for concordance with other variables (e.g. a high score on one measure of depression will be checked with a high score on another measure of depression), if they are different then it will be noted. We will also examine for influential observations in the main analysis models, and examine any such influential observations for outliers. All outliers detected will be taken into account during sensitivity analysis.

3.5. Visit windows

Questionnaires at baseline, 12 months follow up and 24 months follow up will need to be returned within 12 weeks of initial date of questionnaire collection, otherwise they will not be included with in the study. Any questionnaires returned outside this window will be recorded as missing.

3.6. Definition

Data are missing at random (MAR) if, conditional upon the independent variables in the analytic model, the missing-ness depends on the observed outcomes of the variable being analysed but does not depend on the unobserved outcomes of the variable being analysed.

4. DESCRIPTION OF PARTICIPANT CHARACTERISTICS

4.1. Disposition

A flow of clusters through the trial will be summarised in a CONSORT diagram as appropriate for cluster trials[9] that will include the eligibility, reasons for exclusion, numbers randomised to the two treatment groups, losses to follow up and the numbers analysed.

4.2. Baseline characteristics

Continuous data that are approximately normally distributed will be summarised in terms of the mean and standard deviation. Skewed data will be presented in terms of the medians and inter-quartile range. Categorical data will be summarised in terms of frequency counts and percentages. We will summarise all variables by trial arm, both cluster level and individual-level summary data will be provided by table 1, 2 and 3 (Section 8). We will also compare all variables by complete case status (i.e. between those who have all variables measured and those who have some missing outcomes).

5. ASSESSMENT OF STUDY QUALITY

5.1. Eligibility checks

The numbers of teachers and students completing the questionnaires, and reasons for exclusion, will be described.

5.2. Study completion

The final follow up is the questionnaire at T2, 24 months after baseline data collection.

5.3. Protocol deviations

5.4. Specify & justify changes made to the planned statistical analyses

6. ANALYSIS OF EFFECTIVENESS

6.1. Summary of primary and secondary outcomes

Primary outcome summary will be presented as given in table 5 (section 8), while the summary for secondary outcomes will be presented as shown in table 6 (section 8).

6.2. Primary analysis

The tested null hypothesis is that there is no difference in level of teacher wellbeing over the period of 2 years (i.e. repeated measures analysis, no primary timepoint) in schools with peer supporters that have MHFA training and those in schools that do not have such support. The primary analysis will be carried out under the intention to treat principle, analysing participants as randomised without the imputation of missing data.

1. The primary analysis: Repeated measures (random effects) models will be used to examine pattern of change in primary outcome over baseline, T1 and T2, adjusted for stratification variables, sex and years of experience. This will include a random effect for individual participants, and another for school. If models fail to converge, we will use robust standard errors to account for clustering at one of the levels. This model will include every teacher who has at least one measure of the outcome (i.e. at baseline, one or two year follow up). Using a maximum likelihood estimator this analysis is robust to data which are MAR [8, 9].
2. Results will be presented as mean difference in the primary outcome between the trial arms over the follow up period, with associated 95% confidence interval and p- value.

Assumptions and Links for each model:

The link function for this model will be the identity function, and the different model assumptions will be checked to ensure that the model is a good fit for the data.

This will include, check for normality assumptions at the different levels, making sure that homoscedasticity holds, and checking that the assumption of linearity of change holds. If these assumptions do not hold, we will investigate transformations of the data to achieve normality. If the model does not converge, we will use a simpler random effects model accounting for clustering by school only.

6.3. Secondary analyses

The primary analysis (i.e. mixed effects repeated measures model including every teacher who has at least one measure of the outcome i.e. at baseline, one or two year follow up) will be repeated with a treatment by time interaction term added to the model. This will allow estimation of treatment effect at 12 and 24 month follow ups separately.

Analysis of secondary outcomes will include linear, ordinal and logistic regression models dependent on the nature of the outcome variable being analysed (continuous, ordinal or binary respectively).

For secondary individual level outcomes that are measured at baseline, 12 and 24 months (e.g. PHQ-8), repeated measures models will be used, these models will include random effects for clustering by individual (due to repeated measure) and by school. All individuals with at least one observation of the outcome measure will be included in the model for that outcome measure using maximum likelihood under a Missing at Random (MAR) assumption. For each secondary individual level outcome that are measured at baseline, 12 and 24 months, 3 models will be presented:

- Model 1: Unadjusted model – Repeated measures of the outcome regressed on treatment arm accounting for clustering due to repeated measures and by school (using random effects).
- Model 2: Partially adjusted model – Model 1 plus adjustment for stratification variables.
- Model 3: Fully adjusted model – Model 2 plus additional adjustment for covariates. For teacher based outcomes the covariates adjusted for will be sex and years of experience, for student based outcomes sex and ethnicity will be included as covariates.

For each secondary individual level outcome measured only at baseline and 24 months, 3 models will be presented:

- Model 1: Unadjusted model – Outcome at 24 months regressed on treatment arm and baseline value of the outcome, accounting for clustering by school (using a random effect).
- Model 2: Partially adjusted model – Model 1 plus adjustment for stratification variables.
- Model 3: Fully adjusted model – Model 2 plus additional adjustment for covariates. For teacher based outcomes the covariates adjusted for will be sex and years of experience, for student based outcomes sex and ethnicity will be included as covariates.

For each secondary school level outcome measured only at baseline and 24 months, 2 models will be presented:

- Model 1: Unadjusted model – Outcome at 24 months regressed on treatment arm and baseline value of the outcome (no need for any adjustment for clustering).
- Model 2: Adjusted model – Model 1 plus adjustment for stratification variables.

6.4. Sensitivity analysis

The sensitivity of the results to the assumptions about the missing data will be assessed using multiple imputation for missing outcome data. This will include imputing under a range of assumptions about the missing data mechanism – allowing the missing data mechanism to depend on the unobserved outcome data, the arm of the trial, and both. We will assess the impact of non-response and missing data on teacher WEMWBS and PHQ-8 outcomes and student WEMWBS and SDQ outcomes. The sensitivity of conclusions to late returned questionnaires will also be checked, by repeating the analyses with only those questionnaires returned within 2 days of the initial data collection date. We will also re-run the primary analyses without identified outliers, as well as those without major protocol violation to assess the effect of outliers and protocol violation on the main outcome.

6.5. Subgroup analysis

The effect of the intervention on teacher WEMWBS and PHQ-8 score at follow-up will be tested for interaction with: baseline wellbeing / depression score (grouped as above or below the bottom quartile of the WEMWBS, or with a score of 10 or more on the PHQ-8), geographical area (Wales / England), school-level FSM and gender of teacher. The effect of the intervention on student WEMWBS and SDQ score at follow-up will be tested for interaction with baseline wellbeing / SDQ score (grouped as above or below the bottom quartile of the WEMWBS and a score of 16 or more on the SDQ), geographical area, school and individual level FSM, gender and ethnicity. Results will be interpreted with caution due to the low power and number of interactions being tested.

We will use a Complier Average Causal Effect (CACE) approach (using Instrumental Variable analysis or Principal Stratification) [7] to examine the impact of MHFA training on follow up WEMWBS and PHQ-8 scores, comparing those who completed the 1 or 2 days training in the intervention schools with those in the control schools who would have completed the training, had they been offered it. This will be based on matching their gender, role in the school and years of experience with those who were trained. For each participant we will calculate a summary score of the two outcomes, and use the summary score to account for the repeated measurements (thus reducing the 2 measures to

a single measure) and, if only a single measure is available then we will use just the single measure. We will include robust standard errors to account for clustering at the school level.

6.6. Exploratory analysis

Assessing the hypothesised mechanisms of change and influence of school context

In keeping with the study's logic model and hypothesised mechanisms of change, logistic regression models will be used to compare binary measures of stress and satisfaction at work, support given / received at school, school's perceived attitude to staff and student wellbeing and perceived quality of relationships in school between arms at follow up, adjusted for baseline scores, school-level FSM and geographical area. We will also examine whether baseline measures of these variables – which provide indicators of school psychosocial context - moderate the effect of the intervention by including interaction terms between these baseline variables and intervention arm in the analysis model.

Appropriate mixed models will be applied as for the primary analysis (see section 6.2, paragraph 1), but additionally including measures of these mechanisms as covariates. We will assess the degree to which the estimated treatment effect attenuates compared with our chief analysis model – substantial attenuation would indicate the proposed mechanisms of change are indeed acting as such. However, this is subject to accurate measurement of the proposed mechanism of change and the lack of confounding variables influencing the relationship between mechanism and outcome [12].

Assessing the impact of differing levels of implementation on effect of the intervention

Data from the Process Evaluation (from training participant evaluation forms, peer supporter feedback meetings, peer supporter logs of support and follow up questionnaires) will be examined to see if intervention schools can be divided into 'low' or 'high' implementation groups. This exploratory work to categorise schools into groups based on implementation will be carried out blinded to the outcome. Factors, measured as binary variables, that will be considered for this analysis will be:

Dosage

- At least 8% teachers completed the MHFA for Schools training versus less than 8%
- At least 8% of staff completed the MHFA for adults training and went on to become a peer supporter versus less than 8%
- At least 8% of staff still acting as peer supporters by time 2 follow up versus less than 8%

Reach

- Whether 75% or more of teachers attended the one hour awareness training versus less than 75%
- Higher than study mean for those who select 'staff peer supporter' in response to the question 'if a work related problem was making you stressed or down who would you talk to about it at school?' in the follow up questionnaires

Fidelity

- Whether 100% course attendees indicated that all topics were covered for one and two day courses versus less than 100%
- Whether the peer supporters set up a confidentiality policy for the service versus no policy
- Whether the peer support service was advertised in three or more ways initially versus advertised in two or fewer ways
- Whether the peer support service was re-advertised at the beginning of the 2017-2018 academic year versus not re-advertised
- Whether the peer support service has been championed by a member of the senior leadership team versus not championed

Factors relating to quality of the training were considered for inclusion – as measured by training participant feedback forms - but analysis indicated that there was not sufficient differentiation between schools for this to be useful as a high or low implementation indicator.

We will examine patterns of correlation across these different factors, to determine if one composite binary measure low/high implementer can be derived. If correlation is poor and a composite measure would not clearly differentiate high and low implementing schools, correlation among the factors within each sub-heading will be considered, to see if we can determine schools that are high/low dosage, high/low reach and/or high/low fidelity. If this is also not possible then we will conclude that the concept high / low implementers does not reflect the mixed pattern of implementation within each school and we will use qualitative analyses only to consider the impact of the different aspects of implementation (dosage, reach and fidelity) on effectiveness .

If we are able to divide the schools into high / low implementers, we will compare primary outcomes among schools receiving the intervention who are high implementers, with those who are low implementers. Adjustment will be made for both school level and individual level covariates, as in the primary analysis (see section 6.2, paragraph 1). This will be carried out only for those measures of implementation which applied to between 25 and 75% of schools in the intervention arm.

7. ANALYSIS OF SAFETY

School contacts and those delivering the intervention (MHFA trainers, HSCs and those trained as peer supporters) will be asked to contact the study team within 2 working days if any untoward incident or adverse event (AE) occurs to a member of staff or student i) as a direct result of taking part in the WISE study or ii) because of changes that have occurred in the school environment due to participation in WISE (e.g. heightened awareness among staff of mental health problems leading to inappropriate referrals to specialist help sources for 'normal' student behaviour). In these cases, study specific adverse event/incident forms will be completed, recording information on the event. Members of the study teams in Bristol and Cardiff will also be required to complete a form about any incidents or AEs that they encounter during data collections. All adverse event/incident report forms will be discussed with the Principal Investigator to assess seriousness and to explore causality. All AEs deemed to be 'serious' (SAE) will be reported to the Sponsor within 24 hours. Where the SAE is suspected to be related to the intervention and unexpected, the Chair of the TSC and the FHSREC will be notified in writing within 15 days of the study team receiving the initial report. An SAE which is not deemed to be related to the research will be reported to the TSC at the next scheduled meeting.

8. FINAL REPORT TABLES AND FIGURES

8.1. Subject characteristics and background summaries

Table 1. Baseline Characteristics for Teachers

	Control			Intervention		
Variables (Teacher)	**N**	**Mean/%**	**SD**	**N**	**Mean/%**	**SD**
Male (%)						
WEMWBS						
PHQ-8 score						
Presenteeism Score						
Average days Absent past 4 weeks						
Absence (ever/never)						
Previous mental health problem						
Previous MH training • Yes • No • Can't remember						
Job satisfaction (%) • Very satisfied • Satisfied • A little dissatisfied • Dissatisfied • Highly dissatisfied						
Job stressfulness (%) • Not at all stressful • Mildly stressful • Moderately stressful • Very stressful • Extremely stressful						
Experience with school environment • < 1 year • 1-2 years • 3-5 years • 6-10 years • > 10 years						
Ethnicity (%): • White • Mixed • Asian or Asian British • Black or black British • Chinese						
Wanted to seek help from colleague due to stress (%) • Never • Once or twice • Once a term • Once or twice a month • Once or twice a week • Every day • Haven't felt stressed or down						
Good relationship between students and teachers (%) • Strongly agree						

• Agree • Unsure • Disagree • Strongly disagree						
Good relationship between staff (%) • Strongly agree • Agree • Unsure • Disagree • Strongly disagree						

Table 2. Baseline Characteristics for students

	Control			**Intervention**		
Variables (Students)	**N**	**Mean/%**	**SD**	**N**	**Mean/%**	**SD**
Male						
SDQ score						
Student WEMWBS score						
In past year, regularity of going to a teacher for help with social/personal problem • Never • Once or twice • Once a term • Once a month • More than once a month • I haven't had any problems						
Regularity of wanting to ask for help from teacher but unable to in last year • Never • Once or twice • Once a term • Once a month • I hadn't had any problems						
Good relationship between students and teachers • Strongly agree • Agree • Disagree • Strongly disagree						
Ethnicity (%): • White • Mixed • Asian or Asian British • Black or black British • Chinese						

Table 3. Baseline Characteristics for School

	Control			**Intervention**		
School level variables:	**N**	**Mean/%**	**SD**	**N**	**Mean/%**	**SD**
Teacher student ratio						
Teacher retirement and resignation						
Number of teachers						
FSM (%)						
Number of students						
Teacher absence						

Student absence						
Student attainment						

8.2. Study quality summaries

Table 4. Consort Chart

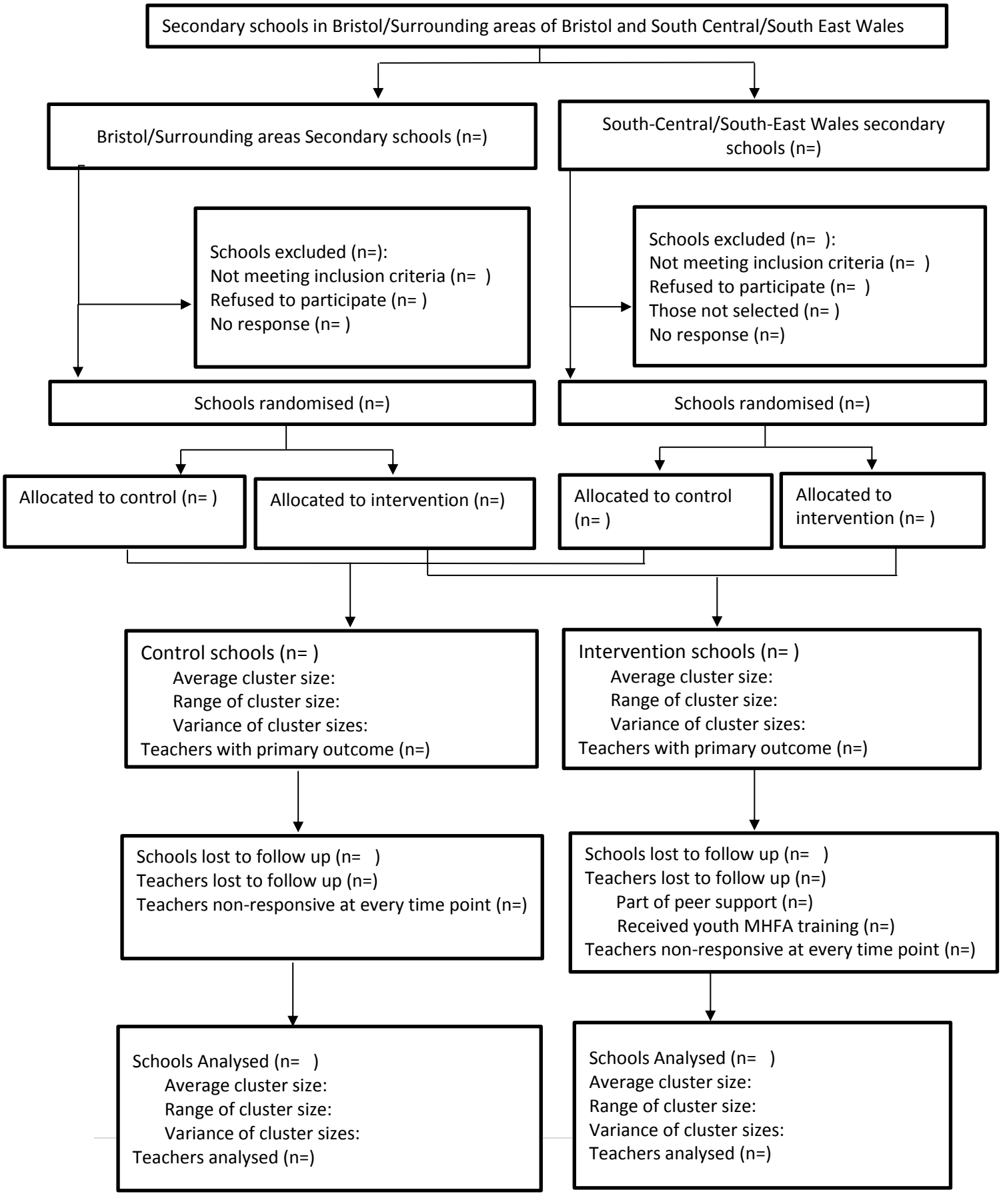


Table 5. Primary endpoint summary

	Control			Intervention		
Variables:	**N**	**Mean/%**	**SD**	**N**	**Mean/%**	**SD**
Teacher WEMWBS Score (T1)						
Teacher WEMWBS score (T2)						

Table 6. Secondary endpoints summary

	Control			Intervention		
Variables (Teachers) :	**N**	**Mean/%**	**SD**	**N**	**Mean/%**	**SD**
PHQ-8 Score (T2)						
Average days absent past 4 weeks (T2)						
Presenteeism Score (T2)						
Absence (ever/never) (T2)						
Teacher absence (school level)						

	Control			Intervention		
Variables (Students) :	**N**	**Mean/%**	**SD**	**N**	**Mean/%**	**SD**
SDQ score						
Student WEMWBS score						
Student absence (school level)						
Attainment (school level)						

8.4. Primary outcome results

Table 7. Primary outcome table

LMM	I vs C[†] adjusted difference in means (95% CI)	p (adjusted difference)
WEMWBS		

[†] I= intervention; C= control.

8.5. Secondary outcomes results

Table 8. Secondary outcomes table

Continuous variables:

	I vs C† unadjusted difference in means (95% CI)	P - value (unadjusted difference)	I vs C† adjusted difference in means (95% CI)	P - value (adjusted difference)
Teacher WEMWBS (T1)				
Teacher WEMWBS (T2)				
PHQ-8 (T2)				
Student WEMWBS				
Student SDQ				

LMM	I vs C† adjusted difference in means (95% CI)	p (adjusted difference)
PHQ-8		

† I= intervention; C= control.

	I vs C† unadjusted odds ratio (95% CI)	P – value (unadjusted difference)	I vs C† adjusted odd ratio (95% CI)	P - value (adjusted difference)
PHQ-8				
Teacher absence[1]				
Teacher presenteeism[1]				

LMM	I vs C† adjusted odds ratio (95% CI)	p (adjusted difference)
PHQ-8[1]		
Teacher absence[1]		
Teacher presenteeism[1]		

† I= intervention; C= control.

[1] Binary outcome

	I vs C† unadjusted odds ratio (95% CI)	P – value (unadjusted difference)	I vs C† adjusted odds ratios (95% CI)	P - value (adjusted difference)
PHQ-8[2]				
Teacher absence[2]				
Teacher presenteeism[2]				

LMM	I vs C† adjusted odds ratio (95% CI)	p (adjusted difference)
PHQ-8[2]		
Teacher absence[2]		
Teacher presenteeism[2]		

† I= intervention; C= control.

[2] ordinal outcome

	I vs C† unadjusted difference in means (95% CI)	P - value (adjusted difference)	I vs C† adjusted difference in means (95% CI)	P - Value (adjusted difference)
Teacher absence[3]				
Student absence[3]				
Student Attainment[3]				

[3] school level outcome

8.6. Sensitivity analysis for primary endpoint

Table 9. Sensitivity analysis primary endpoint results

	Control		Intervention			
	Mean	SD	Mean	SD	I vs C[††]adjusted difference in means (95% CI)	P – Value (adjusted difference)
Excluding. Late replies: Per-protocol						
Outliers Included						
Imputing missing data						

[†] I= intervention; C= control.

8.7. Sensitivity analysis for secondary endpoints

Table 10. Sensitivity analysis secondary endpoint results

Imputing missing data	Control		Intervention			
	Mean	SD	Mean	SD	I vs C[††]adjusted difference in means (95% CI)	P – value (adjusted difference)
PHQ-8						
Student WEMWBS						
Student SDQ						

[†] I= intervention; C= control.

8.8. Subgroup analysis

Table 11. Subgroup analysis results

Teachers WEMWBS		
	I vs C† adjusted difference in means (95% CI)	Bonferroni p-value (adjusted difference)
Baseline WEMWBS*		
Geographical location		
School level FSM		
Gender		

† I= intervention; C= control.

* Binary outcome indicating grouped as above bottom quantile of the WEMWBS

Teachers PHQ-8 score with following interactions:		
	I vs C† adjusted difference in means (95% CI)	Bonferroni p-value (adjusted difference)
Baseline PHQ-8*		
Geographical location		
School level FSM		
Gender		

† I= intervention; C= control.

* Binary outcome indicating a score of 10 or more on the PHQ-8

Student WEMWBS score with following interactions:		
	I vs C† adjusted difference in means (95% CI)	Bonferroni p-value (adjusted difference)
Baseline WEMWBS*		
Geographical location		
Individual FSM		
School FSM		
Gender		

† I= intervention; C= control.

* Binary outcome indicating grouped as above bottom quartile of the WEMWBS

Student SDQ score with following interactions:		
	I vs C† adjusted difference in means (95% CI)	Bonferroni p-value (adjusted difference)
Baseline SDQ*		
Geographical location		
Individual FSM		
School FSM		
Gender		

† I= intervention; C= control.

* Binary outcome indicating a score of 16 or more on the SDQ

8.9. Safety results

Table 12. Adverse events

Adverse event	Serious Adverse event	Intervention	Control

9. APPENDICES

9.1. Stata code for derived variables

9.2. Stata code for final analyses

9.3. Details of standard assessment tools

Include all details of any questionnaires or other assessment tools used in the study.

1. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, Parkinson J, Secker J, Stewart-Brown S. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health and Quality of life Outcomes. 2007 Nov 27;5(1):1.
2. Kroenke K, Strine TW, Spritzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord. 2009; 114(1-3):163-73.
3. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993 Nov 1;4(5):353-65.
4. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. The British Journal of Psychiatry. 2000 Dec 1;177(6):534-9.

10. REFERENCES

1. Kidger J, Brockman R, Tilling K, Campbell R, Ford T, Araya R, King M, Gunnel D. Teacher's wellbeing and depressive symptoms, and associated risk factors: a large cross sectional study in English secondary schools. 2016; 192:76-82.
2. Eldridge SM, Costelloe CE, Kahan BC, Lancaster GA, Kerry SM. How big should the pilot study for my cluster randomised trial be? Statistical Methods in Medical Research. 2016; 25(3):1039-1056.
3. Caille A, Kerry S, Tavernier E, Leyrat C, Eldridge S, Giraudeau B. Timeline cluster: a graphical tool to identify risk of bias in cluster randomised trials. BMJ. 2016; 354; 4291.
4. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, Parkinson J, Secker J, Stewart-Brown S. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health and Quality of life Outcomes. 2007 Nov 27;5(1):1.
5. Kroenke K, Strine TW, Spritzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord. 2009; 114(1-3):163-73.
6. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993 Nov 1;4(5):353-65.
7. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. The British Journal of Psychiatry. 2000 Dec 1;177(6):534-9.

8. Shrier I, Steele RJ, Verhagen E et al.: Beyond intention to treat *Clinical Trials* 2014; 11(1):28-37.
9. Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. Journal of clinical epidemiology. 2013 Sep 30;66(9):1022-8.
10. Cnaan A, Laird NM, Slasor P. Tutorial in biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Stat Med. 1997;16:2349-80.
11. Campbell MK, Piaggio G, Elbourne DR, Altman DG for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012;345:e5661
12. Dunn G, Emsley R, Liu H, Landau S, Green J, White I, et al. Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme. Health Technol Assess 2015;19(93)