EFFECTIVENESS AND COST-EFFECTIVENESS OF HUMANISTIC COUNSELLING IN SCHOOLS FOR YOUNG PEOPLE WITH EMOTIONAL DISTRESS (ETHOS): A RANDOMIZED CONTROLLED SUPERIORITY TRIAL

STATISTICAL ANALYSIS PLAN
FINAL (4th draft): 24/10/2018

Short title	Effectiveness and cost-effectiveness of humanistic counselling in schools for young people with emotional distress (ETHOS)
Lead organization	University of Roehampton
Funding body	Economic and Social Research Council
ISRCTN Number	ISRCTN10460622

Authored by

_____ 29/10/2018
David Saxon, University of Sheffield Date
Principal Statistician

Approved by

_____ __02_/11_/_2018_
Jason Madan, ETHOS Data Monitoring Date
and Ethics Committee Chair

_____ 25/10/2018
Derek Bolton, ETHOS Trial Steering Date
Committee Chair

_____ 22/10/2018
Mick Cooper, University of Roehampton, Date
Principal Investigator

Table of contents

List of abbreviations

AE	Adverse event
AIC	Akaike information criteria
BACP	British Association for Counselling and Psychotherapy
CHI-ESQ	Experience of Service Questionnaire
CI	Confidence Interval
CONSORT	Consolidated Standards Of Reporting Trials
CP	Conduct Problems
CRF	Case Report Form
DMEC	Data Monitoring and Ethics Committee
ESRC	Economic and Social Research Council
ETHOS	Effectiveness and Cost-effectiveness Trial of School-based Humanistic Counselling
FU	Follow-up
GCP	Good Clinical Practice
GBORS	Goal Based Outcome Record Sheet
HA	Hyperactivity
ICC	Intra-cluster correlation
ITT	Intention to treat
LOCF	Last observation carried forward
MAHSC-CTU	Manchester Academic Health Science Centre-Clinical Trials Unit
PCAU	Pastoral Care as Usual
PI	Principal Investigator
PM	Project Manager
PP	Peer problems
PS	Prosocial behaviour
RCADS	Revised Children's Anxiety and Depression Scale
RSES	Rosenberg Self-esteem Scale
SAE	Serious Adverse Event
SBHC	School-based Humanistic Counselling
SD	Standard deviation
SDQ	Strengths and Difficulties Questionnaire
SDQ-ES	Strengths and Difficulties Questionnaire Emotional Symptoms subscale
SES-BE	Student Engagement Scale - Behavioral Engagement subscale
SMD	Standard mean difference
SQL	Standard Query Language
SRA	Social Research Association
TMG	Trial Management Group
TSC	Trial Steering Committee
WEMWBS	Warwick Edinburgh Mental Well-being Scale
WMD	Weighted mean difference
YP-CORE	Young Person's Clinical Outcomes in Routine Evaluation

2 Introduction and Key Trial Objectives

1.1 Study rationale

Pilot studies indicate that a trial of school-based humanistic counselling (SBHC) for young people (aged 13-16 years old) experiencing emotional distress is feasible and that there are initial indications of a short-term effect (e.g., Pearce *et al*., 2017; McArthur *et al*., 2013). However, a trial powered to detect clinically meaningful differences is required which can provide more comprehensive data on the effectiveness of SBHC, in particular its longer-term effects, cost-effectiveness and impact on educational outcomes; as well as identifying mechanisms of change.

1.2 Objectives

The trial will investigate the following main objective:

- To determine whether the addition of SBHC to PCAU leads to greater reductions in psychological distress as compared with PCAU alone in young people with emotional symptoms.

Secondary objectives are to evaluate the effectiveness of SBHC as compared to PCAU on a range of additional outcomes, including depression, anxiety, self-esteem, personal goals, well-being and educational engagement. In addition, we aim to identify the mechanisms of change in SBHC.

This statistical analysis plan is written in conjunction with the International Conference on Harmonisation topic E9 (Statistical principles for clinical trials, 1998), and the published trial protocol (Stafford *et al.*, 2018). The trial will be conducted in accordance with Good Clinical Practice (GCP) in Clinical Trials (International Conference on Harmonisation, 1996).

Planned analyses of costs and other economic analyses will be included in a separate document.

2 Study Design

2.1 Study outline

The ETHOS study is a superiority randomised controlled trial comparing the effectiveness of school-based humanistic counselling (SBHC) with pastoral care as usual (PCAU).

2.2 Study outcomes

Primary outcome:

Emotional distress severity and symptomology as measured by the Young Person's Clinical Outcomes in Routine Evaluation (YP-CORE) (Twigg *et al.*, 2009) at 12 weeks post-randomisation controlling for baseline YP-CORE score.

The YP-CORE is a 10-item, self-report, 5-point Likert-type scale measuring psychological distress in the study population. Participants are asked to rate how they have been feeling over the last week (prior to completing the questionnaire) in relation to 10 items. Individual item scores range from 0 ('not at all') to 4 ('most or all

of the time') with a total YP-CORE score ranging from 0 to 40. The YP-CORE is a clinically relevant measure for assessing changes in psychological distress in the age group being studied and has demonstrated good internal reliability (α=0.80) and test-retest reliability across one week (r=76, 95% CI 0.65 to 0.86) (Twigg *et al.*, 2009).

Secondary outcomes:

a) YP-CORE at 6 weeks and 24 weeks post-randomisation, controlling for baseline
b) The following will also be measured at 6, 12, and 24 weeks, controlling for baseline:
    i. The Revised Children's Anxiety and Depression Scale–Short Version (RCADS-SV) (Ebesutani *et al.*, 2012)
    ii. The Rosenberg Self-esteem Scale (RSES) (Rosenberg, 1965)
    iii. The Student Engagement Scale–Behavioral Engagement subscale (SES-BE) (Lam *et al.*, 2014)
    iv. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS) (Tennant *et al.*, 2007)
    v. The Strengths and Difficulties Questionnaire (SDQ and SDQ follow-up): Total Difficulties (Goodman, 2001)
    vi. Subscales of SDQ: Emotional Symptoms, Peer Problems, Hyperactivity, Prosocial (Goodman, 2001)
    vii. The Goal Based Outcome Record Sheet (GBORS) (Law *et al.*, 2015)
c) The Experience of Service Questionnaire (CHI-ESQ) (Attride-Stirling, 2002) at 12 weeks, to measure satisfaction with treatment provision.
d) Attendance, exclusion, and attainment at 24 weeks, controlling for baseline.

2.3 Randomisation

Consenting young people are allocated to one of two groups, SBHC or PCAU, via remote access to the central randomisation procedure that is hosted by an in-house vb.net application with a SQL server at the Clinical Trials Unit, Manchester Academic Health Science Centre, University of Manchester (MAHSC-CTU). Sequence generation is concealed from both the assessor and young person, as well as the rest of the core research team.

There is no evidence to indicate factors strongly associated with outcomes, therefore the only stratification was by school. The randomisation ratio was 1:1 using the method of permuted blocks within school strata with adjacent block sizes themselves varying randomly within pre-specified limits.

2.4 Blinding

Blinding of participants, providers, and assessors is not possible, so this trial employs an 'assessor blind' design wherein 'Testers'/'Assessors' (i.e., research assistants administering 6 week, 12 week, and 24 week tests) are blind to the young person's allocation for the duration of the trial. To help ensure the success of the

blind, a different Tester will be employed at midpoint, endpoint, and follow-up for each participant. The success of blinding will be assessed by asking Testers to indicate what group they believe the participant they assessed is allocated to on a predesigned Case Report Form (CRF) developed for the trial. In addition, participants will be asked not to reveal the group to which they have been assigned, as far as is possible. We will report the proportion of Testers who believed they knew which group the young person was allocated to, and the proportion of those that correctly predicted the allocation.

Statisticians conducting the analysis will not be involved in the administration of the trial and will be blinded to the randomisation for the primary analysis. The primary analysis will be carried out on a dataset containing only those variables required to address the primary question. In this dataset, key variables (i.e. treatment assignment) will be coded as non-identifiable variables in order to minimise potential biasing in analyses.

2.5 Interim analyses, data monitoring committees, etc.

There are no statistical criteria for stopping the trial early and there will be no requirement of an interim analysis.

The following committees have been established:

a) *Trial Steering Committee (TSC)* includes an independent Chair (not involved directly in the trial other than as a member of the TSC), an independent clinician/counselling academic, an independent economist/academic, a representative young person, a representative parent/carer, and a representative educationalist from the Department for Education. In attendance at these meetings will be a representative from the ESRC (funder), the Principal Investigator (PI), the Project Manager (PM), and the Principal Statistician. The role of the TSC is to monitor the scientific integrity of the trial, the scientific validity of the trial protocol, and assessment of the trial quality and conduct, to ensure that the trial is being conducted in accordance with the principles of good clinical and research practice (as per the British Association of Counselling and Psychotherapy [BACP], ESRC and Social Research Association [SRA] guidelines) as well as for the scientific quality of the final trial report. Decisions about the continuation or termination of the trial or substantial amendments to the protocol are the responsibility of the TSC. The Committee's terms of reference, roles and responsibilities are defined in a charter issued by the Trial Management Group (TMG) that were reviewed at the first TSC meeting.
b) *Data Management and Ethics Committee (DMEC)* to review accruing trial data and to assess whether there are any safety issues that should be brought to the participants' attention, whether any safety amendments should be made, or if there are any reasons that the trial should not continue. The DMEC is independent of the TMG and comprises an independent chair, an independent statistician, an independent clinician specialising in counselling

work with young people, and an independent health economist. The PI and TMG are responsible for nominating DMEC members. The DMEC's terms of reference, roles and responsibilities are defined in a charter issued by the TMG. This charter outlines any stopping rules and the frequency of analysis and DMEC meetings during the trial. The DMEC meets in confidence three times, after the data collection start date, over the course of the trial. Open Reports to the DMEC are prepared by the PM, with any agreed trial data provided by Manchester Academic Health Science Centre-Clinical Trials Unit (MAHSC-CTU), prior to the DMEC meeting. Closed reports are prepared by the MAHSC-CTU, with the support of information provided by the PM. The DMEC Chair report their recommendations to the Chair of the TSC and may request additional reports or information if required. This report is submitted to the TMG, and the PI ensures that all actions and recommendations are followed up.

c) *Trial Management Group (TMG)* includes individuals responsible for the day-to-day management of the trial including the PI and all co-researchers and identified collaborators (including the Principal Statistician and the PM). Notwithstanding the legal obligations of the University of Roehampton and the PI, the TMG have operational responsibility for the conduct of the trial including monitoring overall progress to ensure the protocol is adhered to and taking appropriate action to safeguard the participants and the quality of the trial if necessary.

2.6 Sample Size

Sample size was calculated to take account of likely variability between schools and participants lost to follow-up. Currently, there is no estimate of the size of the intra-cluster correlation (ICC) for schools, therefore, as there is usually a single counsellor in a school, previously reported ICCs for counsellors is considered a proxy for the school ICC in the sample size calculation.

Firstly, without either clustering effects or participants lost to follow-up, for 90% power to detect a standardized mean difference (SMD) of 0.5, 86 participants would be required per arm (172 in total). The effect size was determined by pooling findings on the primary outcome from four previous studies and making a conservative estimate (Cooper *et al.*, 2010; McArthur *et al.*, 2013; Pybis *et al.*, 2014; Pearce *et al.*, 2017). The intra-cluster correlation (ICC) for counsellors was estimated from prior data as 0.05 (Cooper *et al.*, 2010; McArthur *et al.*, 2013; Pybis *et al.*, 2014; Pearce *et al.*, 2017).

On average, we estimated that within each school, nine young people could be seen per counsellor; if there is 20% loss to follow-up, this leaves a mean of 7 participants with outcome data. Considering both the ICC and average number of young people leads to a design effect of 1.31, which, when multiplied by the pre-cluster sample size, gives 1.31 × 86 = 113 (rounded up). Hence, after loss to follow-up has taken place, 113/7.2 = 16 (rounded up) practitioners are required per arm. Add 1 (following Hayes et al., 2009) to give 17 counsellors. To find the number

before loss to follow-up, we calculated 17 × 9 = 153 participants required per arm and 153 × 2 = 306 in total.

3. Data Sources, Data Evaluability, and Analysis Population

3.1 Data sources

The following data were collected from participants on our baseline demographics form, and a counsellor-completed "Current View" form:

a) Age
b) School year
c) Gender
d) Ethnicity
e) Disability
f) Problem description
g) Complexity factors
h) Contextual problems
i) Attendance/attainment difficulties

Data from measures and their collection timepoints are presented in Table 1.

Table 1: Timepoints for the collection of measures included in analyses

Measure	**Description and subscores**	**Base**	**Mid (6 weeks)**	**End (12 weeks)**	**F-U (24 weeks)**
YP-CORE	Psychological distress	X	X	X (primary)	X
SDQ/SDQ-FU	Psychological difficulties and strengths. Total difficulties scale (SDQ-TD) consists of Emotional Symptoms (ES), Conduct Problems (CP), Hyperactivity (HA), and Peer Problems (PP). Strengths subscale is Prosocial behaviour (PS)	X	X (FU)	X (FU)	X (FU)
RCADS-SV	Depression, anxiety	X	X	X	X
RSES	Self-esteem	X	X	X	X
SES-BE	Behavioural engagement at school	X	X	X	X
WEMWBS	Well-being	X	X	X	X
GBORS	Personal goals	X	X	X	X
CHI-ESQ	Satisfaction with service			X	
Attainment	Scored standardised by school of educational attainment	X			X

Attendance	Percentage of days attended over the previous school term	X			X
Exclusions	Percentage of days excluded over the previous school term (e.g., for behavioural reasons)	X			X

All data will be entered into a study-specific database and, prior to analysis, the statistician will review the data and raise any questionable data with the study team.

3.2. Study Population

3.2.1 Inclusion criteria

Participants are young people attending one of 18 secondary schools across London who meet all the inclusion criteria at the time of assessment. This includes:

- Aged between 13 and 16 years of age
- Experiencing moderate to severe levels of psychological distress as assessed by a score of ≥5 on the Strengths and Difficulties Questionnaire Emotional Symptoms (SDQ-ES) Scale
- Considered capable of comprehending the forms, with a guide English reading age of 13 years
- Want to participate in counselling, or want to undertake counselling, or want to see a counsellor
- Not currently in receipt of counselling or any other therapeutic intervention that may be impeded through participation in the trial
- Have a school attendance record of at least 85% as assessed by the school.

3.2.2 Exclusion criteria

A young person is ineligible to take part in the study if any of the following criteria are met at the point of assessment:

- Unable to provide informed consent (not 'Gillick competent')
- Parent/carer of the young person has not provided their informed consent
- At risk of serious harm to self or others at the time of assessment
- Planning to leave the school within the academic year
- Unwilling to complete all assessments
- Unwilling to allow sessions to be audio recorded.

3.2.3 Analysis population

The primary analysis will be conducted on an intention to treat (ITT) population comprising all randomised participants. All randomised participants will complete baseline measures. Groups will be compared based on their randomised treatment assignment.

4. Outline of Analyses

The statistical analysis will be implemented in a validated statistical software package (primarily Stata, see Appendix 2). Data will be reported according to the Consolidated Standards of Reporting Trials (CONSORT) statement and the extension for non-pharmacological trials (Boutron *et al.*, 2017).

4.1 General considerations

a) Summaries of continuous variables will comprise the number of observations used, mean, median, standard deviation, inter-quartile range, minimum, and maximum as appropriate for the distributional form of the data.
b) Summaries of categorical variables will comprise the number of observations used, and the number and percentage of observations in each category.
c) Tables containing the results of the statistical modelling will present the overall difference between treatment groups and list each model covariate and its effect with 95% confidence intervals (CI).

Details of data derivations and methods of handling missing data are covered in section 5.

4.2 Disposition

A CONSORT style diagram (template in Appendix 1) will be presented to summarise recruitment of participants into the trial for all participants screened. The diagram will also summarise attendance throughout the study. The following summary will be presented for the ITT population:

a) Non-attendance – the number of sessions attended
b) Early termination of therapy –the number and percentage of participants that withdraw from the study
c) Switch over – the number and percentage of participants that switch treatments arms.

Reasons for early termination and switch over will be summarised.

4.3 Demographics and baseline characteristics

The following summaries of baseline data, as outlined in 4.1 will be presented, overall and for each treatment:

a) Demographics: age, gender, ethnicity, and disability
b) Baseline scores on YP-CORE and other measures (Table 1)
c) Where clinical thresholds and categorical severity levels for measures are available, these will also be summarised. For example, YP-CORE (Twigg *et al.*, 2016); SDQ (http://sdqinfo.org), and RCADS (Manual).

No statistical tests will be conducted to determine statistically significant differences between the baseline characteristics and demographics. However,

imbalances will be noted, descriptively reported, and considered in a sensitivity analysis (Pocock, Assmann, Enos & Kasten, 2002; Senn, 1994).

4.4 Effectiveness

4.4.1 Primary outcome.

The primary outcome will be YP-CORE score at 12 week follow-up, controlling for baseline YP-CORE score. A mixed effects model will be fitted to the data that will include the following independent variables:

a) Treatment group (fixed effect)
b) Baseline YP-CORE (fixed effect)
c) School (random effect).

School is likely to be a source of outcome variability as it will impact similarly on participants in the two allocated treatment groups in the same school and differently to participants in another school. An ICC derived from previous studies was included in the sample size calculation to represent this effect. Therefore, 'school' will be fitted as a random intercept term in the mixed model to adjust for the variability between schools. In most cases, this is considered preferable to treating schools as a fixed effect (Kahan & Morris, 2013). Most schools have only one counsellor and one PCAU practitioner/team, therefore it is not possible to consider both school and practitioner variability separately in the same 3-level model. Secondary analysis will consider practitioner variability.

Model diagnostics will be carried out to qualitatively inspect that the modelling assumptions are valid and that the model fits the data well:

a) Q-Q plots to assess normality of the residuals for random and fixed parts of the model
b) Scatter plots of fitted values and residuals – constant variance
c) Independence – scatter plot of residuals
d) Adequacy of the fixed effects structure of the model - scatterplot

The results of the model will be reported with the parameter estimates, 95% CI, and the *p*-value of all covariates fitted in the model, together with the overall log-likelihood and the AIC for the model. This model will be interpreted fully in the context of the study in order to address the main research question. Standardised effect size, computed using the model estimate difference between groups divided by the baseline pooled SD, will also be reported with 95% CI.

Exploratory analysis will be carried out to summarise the data collected for the primary outcome:

a) Histograms and boxplots of continuous variables where appropriate stratified by treatment group

b) Scatterplots of continuous variables against YP-CORE outcome score
c) Boxplots of categorical variables against the YP-CORE (including school to examine between school variation)
d) Summary statistics for continuous variables by treatment group including mean, median, standard deviation (SD), interquartile range, minimum and maximum.

Secondary analysis will produce models to explore the effects of covariates on the results. The choice of covariates to study will be influenced by the exploratory data analysis. This analysis will consider both prognostic factors (i.e. factors that may affect response to either treatment) and effect modifiers (i.e. factors that may interact with the difference between SBHC and PCAU). These models will include appropriate interaction terms, including cross-level interactions (e.g. between school and treatment). Improvements in model fit will be tested by comparing reductions in the -2*loglikelihood ratio to the chi squared statistic for the additional degrees of freedom.

4.4.2 Secondary outcomes.

The secondary outcomes include:

a) Score on the YP-CORE score at 6 weeks and 24 weeks post-randomisation, adjusting for baseline score.
b) Scores on other outcome measures collected at baseline, at 6 weeks, at 12 weeks, and at 24 weeks (Table 1).
c) Satisfaction with treatment in the SBHC and PCAU groups at 12 weeks.

The same analysis procedures will be used for these secondary outcomes as for the primary outcome and the results will be presented in the same way. Forest plots will be used to display the results for the different outcome measures. In addition:

a) Attainment, attendance, and exclusions between baseline and 24 weeks will be compared using parametric or non-parametric methods, appropriate to the data
b) Where measures provide clinical thresholds and categorical bands, changes between baseline and 6 weeks/12 weeks/24 weeks will be summarised
c) For the sub-group of participants above the clinical threshold on the YP-CORE at baseline (see Twigg *et al.*, 2016), the proportions making statistical reliable and clinical improvement at 12 weeks in each treatment arm will be compared (Jacobson & Truax, 1991). The cutpoints for this is as follows: for statistically reliable change, YP-CORE scores must change by more than 8.3 points (male, 11–13 years), 8.0 points (male, 14–16 years and female, 11–13 years) and 7.4 points (female, 14–16 years); for clinical change, scores must cross the following YP-CORE cut-off points: 10.3 (male, 11–13 years), 14.1 (male, 14–16 years), 14.4 (female, 11–13 years) and 15.9 (female, 14–16 years).

d) Practitioner variability will be considered using a mixed effects model. Appropriate covariates, including treatment arm, will be included and practitioner will be included as a random effect.

5. Missing and Spurious Data

5.1 Data errors

Prior to locking the dataset for analysis, any obvious errors will be checked with the clinical trials unit (e.g., miscoded data), those that remain errors will be classed as missing. Similarly, any outlying data (defined as more than 3 times the interquartile range either above the 3rd or below the 1st quartiles, by group and time point) will be classed as missing. Missing outcome data will be imputed as below. A sensitivity analysis will be run which includes any outlying data.

5.2 Missing items

Questionnaires will be scored using their own guidelines for dealing with individual missing item responses; usually prorating. If item missingness exceeds developer thresholds, the measure will be considered as missing.

5.3 Missing outcomes and imputation

The number of missing YP-CORE scores and other measures at different timepoints will be summarised, with reasons, overall and by treatment. A non-completion rate of over 20% will be highlighted throughout, including the abstract and used as a caveat. Although there is no reason to believe missing outcomes will be missing not at random (MNAR), logistic regression will be used to test baseline predictors, including treatment, for completion of YP-CORE at 12 weeks post randomisation or not. This will inform whether missingness is completely at random (MCAR) or associated with other variables.

As the primary analysis uses the full ITT sample, all those randomised will be included in the primary analysis, therefore missing YP-CORE scores at 12 weeks will be imputed. Although multiple imputation was considered, the absence of data, e.g. sessional data to identify trends, and the expected small number of missing outcomes, resulted in the adoption of last observation carried forward (LOCF). In addition, a sensitivity analysis using worse-case and best-case scenario imputation will be carried out and the results presented (5.3.2). LOCF has been criticised for possibly introducing bias, therefore the primary results will be assessed in the context of the descriptives and analysis of non-completion and of the sensitivity analyses, outlined below.

5.3.1 LOCF

Where a measure was not collected at 12 weeks, the individual’s score from the mid-point (6 weeks) will be used. If this is not available, then the baseline score will be used.

5.3.2 Worse-case and best-case

For this imputation, two datasets will be created, one which represents the worst case scenario for the treatment of interest (SBHC) while the other represents the best case scenario.

1. *Worst-case*: Missing 12 weeks YP-CORE scores in SBHC are imputed with the highest (i.e. worst) outcome score of SBHC participants with the same baseline score. Missing 12 weeks YP-CORE scores in PCAU are imputed with the lowest outcome score of PCAU participants with the same baseline score.
2. *Best-case*: Missing 12 weeks YP-CORE scores in SBHC are imputed with the lowest (i.e. best) outcome score of SBHC participants with the same baseline score. Missing 12 weeks YP-CORE scores in PCAU are imputed with the lowest outcome score of PCAU participants with the same baseline score.

5.4. Sensitivity analysis

Planned sensitivity analyses are:

- Imputations – analysis of the full sample after analysis of worst-case and best case scenarios.
- Completer comparison -- a sample that only includes those participants with a score at 12 weeks follow-up.
- Per-protocol -- a sample that only includes those participants that received treatments as per protocol. This will be defined as: (a) attendance at a minimum of three counselling sessions (50% of the number of sessions considered to constitute an 'acceptable dose', six sessions); and (b) the counsellor is assessed as meeting adherence criteria to SBHC, as assessed by our PCEPS-YP auditing procedure.
- If differences are found between the two treatment groups on any baseline variable these will be included as covariates in an expanded model including all significant variables and interactions.

These analyses will use the same methods as the primary analysis and results will be summarised similarly (with details in Appendices) in order to assess the impact that the different approaches have had. A dummy table summarising the ITT analysis is presented below. Separate tables will summarise the other sensitivity analyses.

Table 2. Table shell for summarising the ITT analysis for the primary outcome (YP-CORE)

	SBHC (N=X)		PCAU (N=X)		Adjusted difference for baseline (95% CI)
	Baseline Mean (SD)	12 week follow-up Mean (SD)	Baseline Mean (SD)	12 week follow-up Mean (SD)	
ITT (LOCF)	X.X (X.X)	X.X (X.X)	X.X (X.X)	X.X (X.X)	X.XX (X.XX, X.XX)
ITT worst case	X.X (X.X)	X.X (X.X)	X.X (X.X)	X.X (X.X)	X.XX (X.XX, X.XX)

ITT best case	X.X (X.X)	X.X (X.X)	X.X (X.X)	X.X (X.X)	X.XX (X.XX, X.XX)

The main text and abstract will summarise the impact of sensitivity analyses if they:

1. Lead to differences in terms of statistical significance; or
2. Would lead to a change in the effect size descriptor (e.g., from Cohen's "medium" to "small" effect).

6. Safety Outcomes

Where a participant switches groups, any AEs occurring after a switch will be denoted as such in the output.

Possible AEs include school exclusion, suicidal intent, and significant increase in emotional difficulties.

It is unlikely that the number of AEs will permit any inferential analysis, therefore the following summaries will be presented for the ITT as summarised in Table 2 below. AEs will be considered overall and by treatment allocation.

a) Serious adverse events (SAE): the number and percentage of participants recorded as experiencing each SAE
b) All AEs: The number and percentage of participants recorded as having any AE recorded, overall and by AE type
c) Description, date, and time to event

Table 3: Table shell for AEs

	SBHC (N=xxx)		PCAU (N=xxx)	
	No. (%) of participants with an event	No. of events (No. after switching)	No. (%) of participants with an event	No. of events (No. after switching)
Any event	nn (xx%)	nn (nn)	nn (xx%)	nn (nn)
SAE Type1	nn (xx%)	nn (nn)	nn (xx%)	nn (nn)
SAE Type2	nn (xx%)	nn (nn)	nn (xx%)	nn (nn)
…….				
AE Type1	nn (xx%)	nn (nn)	nn (xx%)	nn (nn)
AE Type2	nn (xx%)	nn (nn)	nn (xx%)	nn (nn)
…….				

7. Modifications to original protocol analysis statement

This analysis plan provides more detail than the original protocol statement. Notably this plan clarifies the primary outcome and endpoint at 12 weeks follow-up in order to address the primary research question. Only baseline YP-CORE score and YP-CORE score at 12 weeks will be included in the primary analysis. Comparisons at other timepoints are now defined as secondary outcomes. This plan also provides more detail regarding planned sensitivity analysis.

8. References

Attride-Stirling, J. (2002) Development of methods to capture users' views of child and adolescent mental health services in clinical governance reviews. *Commission for Health Improvement.*

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal Of Statistical Software*, *67*(1). http://doi.org/10.18637/jss.v067.i01

Boutron, I., Altman, D.G., Moher, D., Schulz, K.F., Ravaud, P.D.J.C, *et al*. (2017) CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. *Annals of Internal Medicine*. American College of Physicians; Jul 4;167(1):40. PMID: 28630973

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, M., Rowland, N., McArthur, K., Pattison, S., Cromarty, K., *et al.* (2010) Randomised controlled trial of school-based humanistic counselling for emotional distress in young people: feasibility study and preliminary indications of efficacy. *Child Adolesc Psychiatry Ment Health*; 4:1–12

Ebesutani, C., Reise, S.P., Chorpita, B.F., Ale, C., Regan, J. *et al.* (2012) The revised child anxiety and depression scale-short version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychol Assess*.; 24:833-845

Goodman, R. (2016) Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry*; 40:1337-1345.

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

Hayes, R.J. & Moulton, L.H. (2009) Cluster Randomised Trials. Boca Raton, FL: Chapman & Hall/CRC

Jacobson, N. S., & Truax, P. (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology; 59,* 12–19

Kahan, B.C. & Morris, T.P. (2013). Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? (2013) Stat Med. 2013 32(7):1136-49. doi: 10.1002/sim.5667.

Lam, S.F., Jimerson, S., Wong, B.P., Kikas, E., Shin, H. *et al.* (2014) Understanding and measuring student engagement in school: the results of an international study from 12 countries. Sch Psychol Q; 2014;29:213-232.

Law, D., Jacob, J. (2015) Goals and Goal Based Outcomes (GBOs): Some useful information. 3rd ed. London: CAMHS Press at EBPU

McArthur, K., Cooper, M., & Berdondini, L. (2013). School-based humanistic counseling for psychological distress in young people: Pilot randomized controlled trial. Psychotherapy Research, 23(3), 355-365. doi: 10.1080/10503307.2012.726750

Nieuwenhuis, R., Te Grotenhuis, M., & Pelzer, B. (2012). Influence.ME: tools for detecting influential data in mixed effects models. *R Journal*, *4*, 38–47.

Pearce, P., Sewell, R., Cooper, M., Osman, S., Fugard, A. J. B., & Pybis, J. (2017). Effectiveness of school-based humanistic counselling for psychological distress in young people: Pilot randomized controlled trial with follow-up in an ethnically diverse sample. *Psychology and Psychotherapy: Theory, Research and Practice*, *90*, 138–155. http://doi.org/10.1111/papt.12102

Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statistics in Medicine, 21(19), 2917–30. http://doi.org/10.1002/sim.1296

Pybis, J., Cooper, M., Hill, A., Cromarty, K., Levesley, R. *et al.* (2014) Pilot randomised controlled trial of school-based humanistic counselling for psychological distress in young people: outcomes and methodological reflections. *CPR*;15:241–50.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2016). MLwiN (Version 2.36): Centre for Multilevel Modelling, University of Bristol

Rosenberg, M. (1965) Society and the adolescent self-image. Princeton, NJ: Princeton University Press

Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in Medicine*, *13*, 1715–1726.

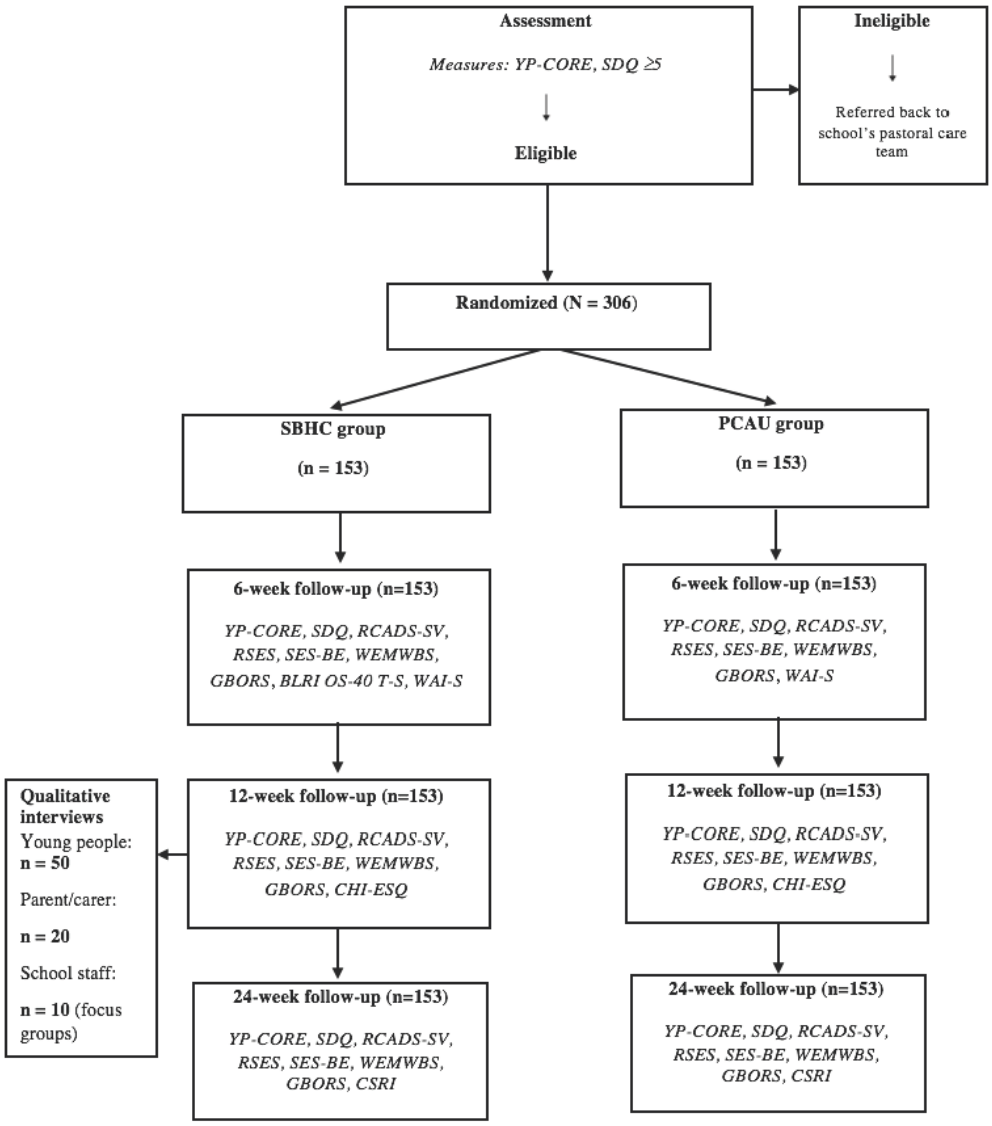
Stafford, M. R., Cooper, M., Barkham, M., Beecham, J., Bower, P., Cromarty, K., … Street, C. (2018). Effectiveness and cost-effectiveness of humanistic counselling in schools for young people with emotional distress (ETHOS): study protocol for a randomised controlled trial. *Trials*, *19*(1), 175. http://doi.org/10.1186/s13063-018-2538-2

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S. *et al.* (2007)The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes;5:63.

Twigg, E., Cooper, M., Evans, C., Freire, E., Mellor-Clark, J., McInnes, B., & Barkham, M. (2016). Acceptability, reliability, referential distributions and sensitivity to change in the Young Person's Clinical Outcomes in Routine Evaluation (YP-CORE) outcome measure: replication and refinement. *Child and Adolescent Mental Health*, *21*(2), 115–123. http://doi.org/10.1111/camh.12128

Appendices

Appendix 1: Template CONSORT diagram



Legend: *YP-CORE* Young Person's Clinical Outcomes in Routine Evaluation; *SDQ* Strengths and Difficulties Questionnaire; *RCADS-SV* Revised Children's Anxiety and Depression Scale - Short Version; *RSES* Rosenberg Self-esteem Scale; *SES-BE* Student Engagement Scale - Behavioural Engagement subscale; *WEMWBS* Warwick- Edinburgh Mental Well-being Scale; *GBORS* Goal Based Outcome Record Sheet; *CHI-ESQ* Experience of Service Questionnaire; *ORS* Outcome Rating Scale; *BLRI OS-40 T-S* Barrett Lennard Relationship Inventory Student form; *WAI-S* Working Alliance Inventory Short Form

Appendix 2: Software

A variety of software can do the analyses and the software used will be specified in write ups. Some candidates are as follows:

- Stata: StataCorp. 2017. Stata Statistical software: Release 15. College Station, TX:StataCorp LLC
- MLwiN (Version 2.36): (Rasbash, J. et al.,2016).