STATISTICAL ANALYSIS PLAN

Solution Focused Brief Therapy (SFBT) in 10–17-year-olds presenting at police custody: A Randomised Controlled Trial with internal pilot

University of Warwick; Centre for Trials Research, Cardiff University

Principal investigators: Dr Samantha Flynn, Professor Peter Langdon

YOUTH ENDOWMENT FUND

YEF statistical analysis plan

Project title	Solution Focused Brief Therapy (SFBT) in 10-17-year-olds presenting at policy custody: A Randomised Controlled Trial with internal pilot
Developer (Institution)	Lancashire and South Cumbria NHS Foundation Trust
Evaluator (Institution)	University of Warwick; Centre for Trials Research, Cardiff University
Principal investigator(s)	Dr Samantha Flynn, Professor Peter Langdon
SAP author(s)	Dr Paul Thompson, Dr Rebecca Playle

Trial design	Two arm parallel randomised controlled trial with internal pilot
Trial type	Effectiveness
Evaluation setting	Community based settings
Target group	Children and young people (CYP) (aged 10-17 years) presenting at a custody suite in Lancashire and South Cumbria NHS Trust and Midland NHS Trust regions, who are referred to the Liaison and Diversion (L&D) team.
Number of participants	282 CYP
Primary outcome and data source	Self-Report Delinquency Measure (SRDM) score
Secondary outcome and data source	1. Criminal offence data-arrest, caution, reprimands, warnings, and conviction data for participants (data held in the Police National Computer).

	2. CYP well-being: the parent/guardian and self-report versions of the Strengths and Difficulties Questionnaire (SDQ) (including internalising, externalising, and prosocial behaviours). 3. Gang Affiliation: The Gang Affiliation Risk Measure (self report)
Potential moderators	• Callous and Unemotional Traits: 24-item Inventory of Callous and Unemotional Traits – Parent/guardian Report and Youth Self-Report Versions. • Learning disabilities (LD): Estimated Verbal IQ based on two subtests of the Wechsler Abbreviated Scale of Intelligence. • Retrospective or in-custody recruitment
Planned number of sites	7 custody suites in Lancashire and South Cumbria, and Midlands and online (eight "sites" in total).
Inclusion criteria	• Aged between 10 to 17 years. Referred to the Liaison & Diversion team by the police.
Exclusion criteria	• A clinician has judged that the child or young person is presenting with a mental illness of a nature and degree warranting immediate intervention from specialist services, including assessment for detention under the Mental Health Act. • The young person is to be remanded into custody. • A child or young person aged 16 years or older judged to lack mental capacity to decide about participating in this trial by staff responsible for gaining informed consent. • The child or young person is unable to converse in English. • Parents/guardians are unable to converse in English (at least one must be able to converse in English to complete parent/guardian measures). • Parents/guardians of under 16s judged to lack mental capacity to decide about participating in this trial by staff responsible for gaining informed consent.
Treatment duration	6 bi-weekly sessions over 12 weeks
Follow-up duration	6 months post-randomisation (single time point for data collection at 6 months post-randomisation)
Planned trial period	36 months
Primary objective	To determine whether there is a difference between support as usual (SAU) plus Solution Focused Brief Therapy (SFBT) and SAU alone in reducing offending behaviours in 10–17-year-olds presenting at a police custody suite.

Secondary objectives	1. Complete an Internal Pilot in the first seven months to examine whether recruitment and randomisation is feasible. 2. Generate evidence to consider whether SFBT + SAU differs from SAU only on externalising and internalising behaviours. 3. Examine whether there is a relationship between changes in externalising and internalising behaviours and changes in offending behaviours. 4. Carry out exploratory sub-group analyses of outcomes by evidence of a learning disability, and callous- unemotional traits. 5. Monitor and report any adverse events. 6. Evaluate whether those excluded from school at the point of enrolment or during the trial will have a different rate of offending behaviour than those who have not been excluded
Intervention	Solution Focused Brief Therapy (SFBT). Six 1-hour sessions will be delivered over 3 months.

SAP version history

Version	Date	Changes made and reason for revision
V0.4	16/05/2023	Minor edits made following comments/suggestions from Dr Izzy Coleman, TSC statistician, University of York Clinical trials unit.
V1.0	18/10/2023	Edits to sample size calculation and randomisation document following inclusion of additional sites.
V1.1	01/11/2023	Minor edits to justifications of sample size following suggestions by TSC statistician, Dr Izzy Coleman.
V2.0	02/02/2024	Primary analysis changed to YEF reviewer suggestion. This was also confirmed by TSC statistician Dr Izzy Coleman and Senior Trial Statistician, Dr Rebecca Playle.

V3.0	20/06/2024	Changes to analysis removing the 12 month followup time point Edit primary analysis to ANCOVA model rather than MLM due to removal of 3rd time point.
V4.0	12/03/2025	"time since randomisation" covariate added to the primary analysis as a binary variable, "in time window" or "out of time window".
V4.1	08/04/2025	Updates to reflect change from stratified block randomisation to minimisation.
V4.2	18/06/2025	Updated to include additional CACE sensitivity analyses at all possible levels of attendance 1-6 sessions. Also corrected a typographical error on the Adherence definition when referring to progression criteria.
V4.3	15/07/2025	Further typographical updates and updated text to reflect recruitment of additional NHS trusts.

ROLES AND RESPONSIBILITIES

Trial Statistician:	Dr Paul Thompson		
Role:	Associate Professor in Applied Statistics (University of Birmingham)		
Date:	15/07/2025	Signature:	
Senior Statistician:	Dr Rebecca Playle		
Role:	Interim Director of Statistics, Centre for Trials Research, Cardiff University		
Date:	15.7.25	Signature:	

Chief Investigator (s):	Dr Samantha Flynn / Professor Peter Langdon		
Role:	Assistant Professor / Professor (University of Warwick/University of Birmingham)		
Date:	15.07.2025	Signature:	
Other non-signatory contributor to the SAP:		Professor Richard Hastings	
Role:	Head of Department, Professor, and Cerebra Chair of Family Research (CIDD, University of Warwick)		

Table of contents

1. Introduction

This statistical analysis plan provides guidelines for the final presentation and analysis for the Solution Focused Brief Therapy for young people in contact with the criminal justice system trial (Solutions). This plan, along with all other documents relating to the analysis of this trial, will be stored in the Statistical Analysis Master File electronically and/or in hard signed copy formats.

2. Background

2.1 Rationale and research question

A systematic review of 38 best evidence studies (Woods et al., 2011) reported that Solution Focused Brief Therapy (SFBT) led to reductions in internalising and externalising behaviour problems in Children and Young People (CYP). In the proposed research, we will conduct a randomised controlled trial with process evaluation and internal pilot (to assess trial feasibility) to evaluate reduction in offending behaviours where CYP presenting at a police custody suite will be randomly allocated to receive Solution Focused Brief Therapy (SFBT) plus Support as Usual (SAU) or SAU alone.

2.2 Objectives

Our primary objective will be to evaluate whether for a sample of 10–17-year-olds presenting at a police custody suite, there is a difference in offending behaviours between participating in the support as usual (SAU) plus Solution Focused Brief Therapy (SFBT) intervention compared to SAU alone, after adjustment for baseline measurements of primary outcome (Self-Report Delinquency Measure) and stratification variables, Verbal IQ and custody suite.

The secondary objectives are to:

- Complete an internal pilot in the first seven months to examine the feasibility of recruitment and randomisation before continuing with the main trial.
- Generate evidence to consider whether there is a difference between SFBT + SAU and SAU alone on scores for SDQ internalising, externalising and prosocial behaviour outcome measures at 6 month follow up time point.
- Examine whether there is a difference between SFBT + SAU and SAU alone in offending behaviours, specifically the numbers of arrests, cautions, reprimands, warnings and convictions at 6 month follow up (adjusted for baseline).

- Examine whether there is a difference between SFBT + SAU and SAU alone on the Gang affiliation measure (T-GARM) at the 6 month follow up.
- Evaluate whether those excluded from school at the point of enrolment or during the trial will have a different rate of offending behaviour than those who have not been excluded.
- Assess the sensitivity of findings under different assumptions with respect to missing data.
- Carry out exploratory sub-group analyses of the primary outcome by Learning Disability (LD) status, retrospective or in-custody recruitment and callous-unemotional traits.
- Monitor and report adverse events related to SFBT.

3. Study Materials

3.1 Trial design

The trial is a two-arm individually randomised controlled trail (RCT) of SFBT plus SAU versus SAU alone, involving CYP (age 10-17 years old) who have presented at one of eight police custody suites in the Lancashire and South Cumbria NHS Trust and Midland NHS Trust regions. The trial involves an internal pilot to be completed at month 6 from the start of the trial, the set-up phase is planned for five months and the pilot phase for seven months (see section 12 for more details). 282 CYP participants will be recruited. Although not included into the primary analysis, there is a potential influence of therapist clustering within the intervention arm. This will be investigated as an additional analysis specified in section 6.4.

Trial design, including number of arms		*Two-arm parallel randomised control trial*
Unit of randomisation		*Individual participant*
Stratification variables (if applicable)		*Custody suite, Verbal IQ*
Primary outcome	variable	Self-reported delinquency
	measure (instrument, scale, source)	Self Report Delinquency Measure at 6-months post-randomisation (SRDM; Smith & McVie, 2003) which is a short measure comprising 15-items pertaining to antisocial behaviours (e.g., burglary, violence). It requires CYP to respond with yes or no with reference to a time-period (6 months). They then report the estimated frequency of the behaviour, and whether they have ever been caught. Each items frequency is scored 0-5, 6-10 is scored 6 and 11+ is scored 11. Minimum score would be

		0 and maximum number of delinquent behaviours would be 165 (15x11). (appendix 1)
Secondary outcome(s)	variable(s)	1. Criminal offences data (appendix 2) 2. Self-reported and parent-reported emotional and behavioural difficulties (appendix 3) 3. Self-reported gang affiliation (appendix 4)
	measure(s) (instrument, scale, source)	1. Criminal offence data-arrest, caution, reprimands, warnings, and conviction data for participants (data held in the Police National Computer). 2. CYP emotional and behavioural difficulties: the parent/guardian and self-report versions of the Strengths and Difficulties Questionnaire (SDQ) (including internalising, externalising, and prosocial behaviours). 3. Gang Affiliation: The Gang Affiliation Risk Measure
Baseline for primary outcome	**variable**	Self Report Delinquency Measure (SRDM)
	measure (instrument, scale, source)	Self Report Delinquency Measure at Baseline (SRDM; Smith & McVie, 2003)
Baseline for secondary outcome	**variable**	• Criminal offences data • Self-reported and parent-reported emotional and behavioural difficulties • Self-reported gang affiliation
	measure (instrument, scale, source)	• Criminal offence data-arrest, caution, reprimands, warnings, and conviction data for participants (data held in the Police National Computer). • CYP emotional and behavioural difficulties: the parent/guardian and self-report versions of the Strengths and Difficulties Questionnaire (SDQ) (including internalising, externalising, and prosocial behaviours). Gang Affiliation: The Gang Affiliation Risk Measure

All participants (in both trial arms) will complete assessments at baseline, 6- months post-randomisation and will be given a choice of how these are completed. These assessments can be completed in a number of ways, face-to-face, online on a website, via telephone, via videoconferencing, or on paper via the post. Assessment will be monitored throughout the

study and a table describing the frequency of assessment method will be presented in the final report.

3.2 Randomisation

Original randomisation strategy: CYP will be randomised on a 1:1 basis to either the intervention or comparator arm using random permuted blocks, stratified by verbal IQ (dichotomised <70, >=70) and custody suite. A more detailed description can be found in the randomisation protocol (v0.8 24/05/2022; Appendix E). Random permuted blocks with varying block size (sizes 2,4,6) are generated using Stata version 17.0 with the 'ralloc' function (Senior Statistician, Playle generated the lists, so that study statistician, Thompson, remains blind to allocations). The list is then uploaded to the Redcap study database and allocations are automated.

New randomisation strategy: Since imbalance has occurred by arm and stratifying variable we will use minimisation to rebalance the randomization. Allocations will be generated using the minimisation module, TMS2. All allocations will still be recorded into the Redcap database, but the Redcap randomisation module will become redundant.

3.3 Sample size

Sample size calculations were conducted using R version 4.1.2 (2021-11-01).

282 CYP participants will be recruited allowing for up to 20% dropout from the trial (N=225). Recruiting this number of CYP, and on the basis of detecting a minimal clinically important difference (MCID) 0.325 (mean difference of 4 points with SD=12.32), assuming a correlation between baseline and follow-up of 0.5 (Borm, Fransen & Lemmens, 2007) and using a two-sided alpha of 0.05, the trial would then be 80% powered. Our assumptions about the minimally detected effect size (MDES) are informed by previous research by the developers of the SRDM measure (Smith, Shute, Flint, McVie, Woodward and McAra, 2001). They report mean and SD in the development samples and based on expertise in our target population have made a conservative adjustment to use a smaller MDES to reflect some level of uncertainty. We have also included the pre-post correlation based on values obtained from the START trial using the same outcome measure and in a similar population of adolescents (Fonagy et al., 2020). Similarly, the drop out rates for the START trial show that our predicted rate of drop out (20%) is reasonable given the START trial reported rates of 15% in the intervention arm and 10% in the control arm.

Initial sample size estimates were calculated at n= 448 assuming 90% power and a more conservative correlation between baseline and followup, r=0.334. Given that recruitment has been more challenging than expected, with much slower rates of recruitment; we revisited the sample size calculation to see, in light of new information on pre-post correlation, whether the sample size could be adjusted to ensure that the trial could still be delivered with the required time frame of the funder. Power was also reduced from 90% to 80% power to reduce sample size requirements but still maintain sufficient power for the trial.

The sample size has been designed to address the primary analysis only. A consideration of sample size adjustment after the pilot was originally planned but after discussion with steering committee and funder it was decided that it was not required following earlier adjustment due to slower than expected recruitment rate.

		Protocol	Randomisation
Minimum Detectable Effect Size (MDES)		0.325	0.325
Pre-test/ post-test correlations	level 1 (participant)	0.5	0.5
	level 2 (cluster)	-	-
Intracluster correlations (ICCs)	level 1 (participant)	-	-
	level 3 (cluster)	-	-
Alpha[1]		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size		-	-
Number of clusters[2]	intervention	-	-

		Protocol	Randomisation
	control	-	-
	total	-	-
Number of participants	intervention	141	141
	control	141	141
	total	282	282

3.4 Framework

The trial protocol states that the RCT is designed, "to determine whether there is a difference between support as usual (SAU) plus Solution Focused Brief Therapy (SFBT) and SAU alone in reducing offending behaviours in 10–17-year-olds presenting at a police custody suite." Therefore, the trial is on the basis of superiority of the support with additional therapy arm of the trial.

3.5 Interim analysis

No planned interim analyses. Target sample size will not be recalculated, regardless of rate of recruitment.

3.6 Timing of final analysis

All outcomes will be analysed collectively and after the database is locked one month following the last 6 month follow up post-randomisation. One month after completion of baseline data collection and data cleaning, the database will be soft locked to new recruitment and only entry of follow-up data will be permitted. At this point, baseline data summary tables will be generated. After the database is locked after 6 month follow up post randomisation, a baseline data table of completers vs non-completers will also be created.

3.7 Timing of outcome assessment

Outcomes (secondary)	Data collection timepoints

	Screening	Baseline	6-month follow-up
CYP wellbeing self-report: self-report version of the Strengths and Difficulties Questionnaire		X	X
CYP wellbeing parent/ guardian-report: parent-report version of the Strengths and Difficulties Questionnaire		X	X
Gang Affiliation Risk Measure		X	X
MODERATOR: Self-report Callous and Unemotional Traits		X	X
Parent/guardian-report Callous and Unemotional Traits		X	X
MODERATOR: Wechsler Abbreviated Scale of Intelligence (vocabulary and similarities subscales)		X	
Parent/guardian-report other therapies received (including pharmacological)		X	X
Criminal offence data: arrest, caution, reprimands, warnings and conviction data (referrers and the police)		X	X

4. Statistical Principles

4.1 Levels of confidence and p-values

All confidence intervals presented will be 95% and two-sided. In addition, all applicable statistical tests will be two-sided and will be performed using a 5% significance level.

4.1.1 Adjustment of multiplicity

The overall type I error rate for testing support as usual (SAU) plus Solution Focused Brief Therapy (SFBT) trial arm over the control arm SAU only for the primary endpoint will be controlled at the 2-sided 0.05 significance level. Secondary analyses will control the family-wise error rate using the Holm method.

The Holm method, in a stepwise way, computes the significance levels depending on the P-value based rank of hypotheses. For the ith ordered hypothesis $H(i)$, the specifically adjusted significance level is computed:

$$\alpha'(i) = \frac{\alpha}{m - i + 1}$$

where m is the number of hypothesis tests.

The observed P value $p(i)$ of hypothesis $H(i)$ is then compared with its corresponding $\alpha'(i)$ for statistical inference; and each hypothesis will be tested in order from the smallest to largest P values ($H(1), \ldots, H(m)$). The comparison will immediately stop when the first $p(i) \geq \alpha'(i)$ is observed $(i = 1, \ldots, m)$ and hence all remaining hypotheses of $H(j)$ ($j = i, \ldots, m$) are directly declared non-significant without requiring individual comparison.

4.2 Adherence and protocol deviations

4.2.1 Definition and assessment of adherence

SFBT attendance/engagement data will be recorded in logs by practitioners, including: start date of CYP engagement with the intervention; number of sessions offered and completed. Six bi-weekly sessions over 12 weeks will be offered and young people should attend all sessions where possible.

The number of sessions delivered will be recorded by practitioners in Session Summary forms and any implementation challenges recorded.

Adherence is defined as: 4+ sessions attended .

4.2.2 Presentation of adherence

The number and % of participants for percentage of scheduled sessions attended will be presented in a table. Results will be provided for the treatment group.

4.2.3 Definition of protocol deviation

Any deviation from the randomised intervention plan as detailed in the protocol will be considered as a protocol deviation.

4.2.4 Presentation of protocol deviation

Prospective, planned deviations or waivers to the protocol will not be allowed, e.g. participants who do not meet the eligibility criteria or restrictions specified in the trial protocol will not be enrolled.

Any accidental protocol deviations will be adequately documented on the relevant forms and reported to the Chief Investigators immediately.

Deviations from the protocol which occur frequently will be addressed immediately and if appropriate will be classified as a serious breach.

The final analysis will also present the proportions of protocol deviations in a table.

4.3 Analysis population

Children and young people (CYP) (aged 10-17 years) presenting at a custody suite in Lancashire and South Cumbria NHS Trust or Midland NHS Trust regions who are referred to the Liaison and Diversion (L&D) team. Participants could be recruited up to 3 months retrospectively to permit capturing full sample requirement. The intention-to-treat population for primary and secondary analyses will include all eligible randomised participants according to the trial arm to which they were randomised irrespective of session attendance. If an ineligible participant is randomised, they will be removed from the dataset and not included in the analysis. The database has several automated eligibility checks before randomisation, so it is unlikely that an ineligible participant will get to the stage of randomisation.

5. Study population

5.1 Screening data

The following summaries will be presented for all screened CYP (overall and by custody suite):

Enrolment: the number of days recruiting, the number of CYP screened, the number of CYP recruited, the number of screened CYP not recruited, and the reason for non-recruitment.

This information will be included in the CONSORT flow diagram (see appendix for template).

5.2 Eligibility

CYP (aged 10-17 years) will be eligible for this study if they present at a custody suite and are referred to L&D Services. CYP who present with current symptoms of severe mental illness (e.g. psychosis) and are judged to require specialist intervention from child and adolescent mental health will be ineligible.

The number of ineligible patients randomised, if any, will be reported, with reasons for ineligibility. Ineligible patients will be removed from the data and not included into the analysis.

5.2.1 Inclusion criteria

- 10-17 years of age
- Referred to the Liaison and Diversion Team by the police.

5.2.2 Exclusion criteria

- A clinician has judged that the child or young person is presenting with a mental illness of a nature and degree warranting immediate intervention from specialist services, including assessment for detention under the Mental Health Act.
- The young person is to be remanded into custody.
- A child or young person aged 16 years or older judged to lack mental capacity to decide about participating in this trial by staff responsible for gaining informed consent.
- The child or young person is unable to converse in English.
- Parents/guardians are unable to converse in English (at least one must be able to converse in English to complete parent/guardian measures and to provide consent for young people under the age of 16 years).
- Parents/guardians of under 16s judged to lack mental capacity to decide about participating in this trial by staff responsible for gaining informed consent.

5.3 Recruitment

A CONSORT flow diagram (appendix A) will be used to summarise the number of CYP who were:

- assessed for eligibility at screening
- eligible at screening
- ineligible at screening*
- eligible and randomised
- eligible but not randomised*
- randomised to each arm
- received the randomised allocation
- did not receive the randomised allocation*
- lost to follow-up*
- discontinued the intervention*
- randomised and included in the primary analysis

- randomised and excluded from the primary analysis*

*reasons will be provided.

5.4 Withdrawal/Follow up

5.4.1 Level of withdrawal

The participants care will not be affected at any time by declining to participate or withdrawing from the trial because they will still receive services as usual. If a participant initially consents but subsequently withdraws from the trial, clear distinction will be made as to what aspect of the trial the participant is withdrawing from. These aspects will be:

- Withdrawal from intervention (SFBT only)
- Partial withdrawal from future follow-up data collection (e.g., some questionnaires, interviews)
- Withdrawal from previously collected data, prior to data analysis
- Withdrawal of participation in PNC data collection.

Participants cannot withdraw from the trial but still receive the intervention, if they withdraw from the trial then they will receive usual services only. All participants will be included in the primary analysis unless they withdraw their consent for the use of their data.

5.4.2 Timing of withdrawal

Participants have the right to withdraw consent for participation in any aspect of the trial at any time.

5.4.3 Reasons for withdrawal

Participants who consent and subsequently withdraw will complete the trial withdrawal form or the withdrawal form will be completed on the participant's behalf by the site staff/ trial team based on information provided by the participant.

Presentation of withdrawal/Loss to follow up

The number and % of participants that have withdrawn/loss to follow up from the study will be presented in a table for all stages. Results will be provided for the treatment group.

5.5 Baseline participant characteristics

5.5.1 List of baseline data

Participants will be screened at site and eligibility will be assessed. Potential participant details will be passed from the trial site to the trial team in Warwick. The trial team will contact

the participant as per their preferred choice of data collection to take consent and complete the baseline data:

- Age (Years)
- Sex/gender
- Who they live with, and if they are being looked after
- Whether they are in school
- Type of school
- School year
- Ethnicity
- If they have left school, whether they are in work, an apprenticeship, training, the armed forces or unemployed.
- If English is their first language
- GP contact details
- Medications and treatments (including talking therapies that are being received), collected at baseline and 6 months
- Baseline outcome measures completed (WASI-II is to be completed with researcher assistance [telephone, teleconferencing, or face-to-face])
    1. CYP wellbeing self-report: self-report version of the Strengths and Difficulties Questionnaire
    2. CYP wellbeing parent/ guardian-report: parent-report version of the Strengths and Difficulties Questionnaire
    3. Gang Affiliation Risk Measure
    4. Self-report Callous and Unemotional Traits
    5. Parent/guardian-report Callous and Unemotional Traits
    6. LD: Wechsler Abbreviated Scale of Intelligence (vocabulary and similarities subscales)
    7. Parent/guardian-report other therapies received (including pharmacological)
    8. Criminal offence data: arrest, caution, reprimands, warnings and conviction data (referrers and the police) from the preceding six months.

5.5.2 Descriptive statistics

Characteristics of each trial arm group will be summarised descriptively, both as randomised and as analysed in the primary analysis.

Categorical data will be summarised by numbers and percentages. Continuous data that follow a normal distribution will be summarised using means and standard deviations while skewed continuous variables will be summarised using medians and inter-quartile ranges.

Histograms and boxplots will be used to check the distribution and possible outliers for continuous variables. Tests of statistical significance will not be undertaken for baseline characteristics (Senn, 1994); rather the clinical importance of any imbalance will be noted. Parent/guardians will respond to two check lists of seven items each, medications and therapies. The data will relate to process evaluation and will be reported descriptively and with summary statistics to inform description of the population.

6. Analysis

6.1 Outcome definitions

6.1.1 Primary outcome(s)

The primary outcome measure for this trial is the Self Report Delinquency Measure at 6-months post-randomisation (SRDM; Smith & McVie, 2003 Appendix 1).

6.1.2 Timing, units, and derivation of primary

Primary outcome is collected at baseline, and 6-months post-randomisation. The SRDM is a derived total score following Smith & McVie (2003) and the units are a relative measure of delinquency. The SDRM is a measure comprising 15-items pertaining to antisocial behaviours (e.g., burglary, violence). It requires CYP to respond with yes or no with reference to a time-period (6 months). They then report the estimated frequency of the behaviour, and whether they have ever been caught. Each items frequency is scored 0-5, 6-10 is scored 6 and 11+ is scored 11. Minimum score would be 0 and maximum number of delinquent behaviours would be 165 (15x11). On this basis, we may have a skewed continuous distribution, so a log transformation may be required after inspection of model residuals. In addition, there may be a number of individuals where this is their first time in a custody unit, so there is a possibility of floor effects depending on the frequency of their delinquent behaviour. A higher number of delinquent behaviours is bad, so a reduction in the outcome indicates an effective treatment. Baseline and 6-month follow up data will only be used in the primary analysis.

6.1.3 List of secondary outcomes

Secondary participant reported outcome measures include:

- Criminal offence data for participants during the previous 6-month time period (data held in the Police National Computer). We aim to initially collect crime data over the 6-month period prior to the randomisation, at the 6 month follow-up. We aim to analyse the following counts individually:

  - number of arrests,
  - number of cautions,

- number of reprimands,
- number of warnings,
- number of convictions

• Emotional and behavioural difficulties: the parent/guardian and self-report versions of the Strengths and Difficulties Questionnaire (SDQ) will be used to assess CYP well-being (including internalising, externalising, and prosocial behaviours). The SDQ is a robust and well-validated measure of behavioural and emotional problems (Deighton et al., 2014Appendix 2); measured over the preceding 6 months. The following subscales will be analysed individually for both parent/guardian and self-report:

- Internalising problems
- Externalising problems
- Prosocial behaviour

• Gang Affiliation: The Gang Affiliation Risk Measure (T-GARM; Raby & Jones, 2016; Raby, Jones, Hulbert, & Stout, 2017Appendix 3) is a 15-item measure of gang affiliation that was developed with teenagers.

6.1.4 Order of testing

Secondary participant reported outcomes are tested in the order listed in section 6.1.3.

6.1.5 Timing, units and derivation of secondaries

Secondary outcomes are generally collected at baseline, and 6-months post-randomisation. See section 3.7. Appendix

• Criminal offence data: We aim to initially collect crime data over the 6-month period prior to the commencement of treatment, at the 6 month follow-up. Baseline, and 6-month follow up data will be collected and used in secondary analysis. For each measure, a count will be recorded.

• Emotional and behavioural difficulties: Baseline and 6-month follow up data will only be used in the secondary analysis. The SDQ consists of 25 items which are each scored on a 3-point Likert scale (0, 1, 2). Three subscales will be used: i) Externalising problems - Ranges from 0-20 and is generated by summing the scores of the conduct and hyperactivity subscales; ii) internalising problems - Ranges from 0-20 and is generated by summing the emotional and peer problems subscales; and iii) prosocial behaviour – ranges from 0-10 and is generated by summing prosocial behaviour items. Total scores for the subscales can be generated if no more than three items are missing, otherwise a missing value is generated for the subscore.

- Gang Affiliation: Baseline and 6-month follow up data will only be collected and used in secondary analysis. There are 15 binary (yes/no) items that are summed giving a range 0-15 total score. The score will be analysed as a continuous measure but to aid interpretation, a total score of 7 or more would indicate risk of gang affiliation and would suggest early intervention support is provided. The measure developers provide no guidance on item level missingness, or scoring with missingness. Strategies for dealing with missing data are detailed in section 6.3 for this measure.

6.2 Analysis methods

6.2.1 List of methods and presentation

Internal Pilot study

Statistical analysis for internal pilot feasibility outcomes will be primarily descriptive. Feasibility outcomes (primary outcome measures and all secondary measures) will be estimated as frequencies and percentages, means and standard deviations, or medians and interquartile ranges as appropriate. Feasibility outcomes will be assessed against the pre-specified progression criteria (see Appendix D).

Primary outcome analysis

Our primary outcome analysis will include all randomised participants who provide outcome data (i.e., a modified intention to treat analysis set) and compare mean scores between arms on the SRDM at 6-months post-randomisation using linear regression, adjusting for baseline SRDM score, Verbal IQ, Sex, Age, and custody suite (include 'online') to investigate the overall effect of the intervention on post-randomisation measures.

$$Y_i = \beta_0 + \beta_1 SDRM_{BL_i} + \beta_2 TX_i + \beta_3 VIQ_i + \beta_4 Age_i + \beta_5 Sex_i + \beta_6 timewindow_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

where, Y_i re the SDRM scores; $SDRM_{BL}$ are the baseline SDRM scores; TX is the treatment/control variable indicator; *Custody* is the indicator of custody suite (strata, 8 levels: Blackburn, Preston, Blackpool, Harrow, Burnley, Hatfield, Stevenage and online. These will be included as fixed effects rather than a random intercept); *VIQ* is the Verbal IQ of the CYP (binary; ≤ 70 or >70); *Age* is the continuous measure of age in years, *Sex* is the biological sex indicator of the adolescent, *timewindow* is a binary indicator to distinguish if time to randomisation was "within time window" or "outside time window", and ε_i is the individual

level variation. Custody suite has been introduced into the model as a fixed effect as it is a stratifying variable in the randomisation (Gelman & Hill, 2006).

We will use simple coding for the contrast of custody suite, so that our intercept retains the grand mean and nominally use "online"as our reference level.

Distributional assumptions for the primary linear model will be checked and alternative methods are listed in section 6.2.4

Secondary outcome analysis

The SDQ for both parent-report and self-report versions (analysed separately) and the T-GARM will be analysed following the same method as the primary outcome. The distributions of these secondary outcomes will be assessed prior to conducting the analysis. If skew is significant and residuals assumptions are not met, then a Poisson or negative Binomial model will be specified (see below, under count variables). If range restriction is apparent (significant floor and ceiling effects in distribution plots), then we will use a Tobit regression (Twisk & Rijmen, 2009;Tobin, 1958), as follows:

$$Y_i = \beta_0 + \beta_1 GARM_i + \beta_2 TX_i + \beta_3 VIQ_i + \beta_4 Age_i + \beta_5 Sex_i + \beta_6 timewindow_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$Y_i = \begin{cases} a & Y_i^* \quad if\ Y_i^* \leq l \\ Y_{ijk}^* & if\ l < Y_i^* < r \\ b & Y_i^* \quad if\ Y_i^* \geq r \end{cases}$$

Where l and r are the left and right censoring thresholds respectively. Y_i^* is considered to be a latent partially observed variable that is able to take values beyond the thresholds.

Remaining secondary outcomes, number of Criminal offences (arrests, cautions, reprimands, warnings, and convictions), will be analysed similarly but use Generalized linear model given that these are counts. For count variables, we will use a Poisson (or negative Binomial, as necessary) model checking for zero inflation and overdispersion, as follow:

$$g(Y_i) = \beta_0 + \beta_1 BL_i + \beta_2 TX_i + \beta_3 VIQ_i + \beta_4 Age_i + \beta_5 Sex_i + \beta_6 timewindow_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

Note: $g(.) = log_e(.)$, where $g(.)$ Is the log link function for the secondary outcome measures, whereas the primary outcome, $g(Y_i) = Y_i$. *BL* is the baseline number of offences.

Effect sizes will be calculated based on the adjusted mean difference between the SAU plus intervention and SAU alone group (controlling for baseline) using the formula (Hedges, 2007),

$$Hedges'g = \frac{M_1 - M_2}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{\frac{(SD_1^2 + SD_1^2)}{2}}$$

The effect size will also be reported with 95% confidence intervals defined,

$$g \pm \Phi^{-1}(1 - (\alpha/2))g_{se}$$

Where Φ^{-1} is the percent point function of the normal distribution, and g_{se} is the standard error of the g statistic.

$$g_{se} = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{g^2}{2(n_1 + n_2)}}$$

All parameter estimates from the models will be reported with 95% confidence intervals.

Effect sizes from count models will report rate ratios derived by exponentiating the parameter estimates.

For the remaining secondary outcomes, their effect sizes will be reported as either Hedges' g (Tobit, same as primary outcome) or rate ratios (all other secondary outcomes, exponentiated parameter estimates), given that generalized linear models with log link function are used to model the data and that the measures are positively scored integers with some amount of skew anticipated (Barnett and Dobson, 2008).

6.2.2 Covariate adjustment

We will assess any imbalance of baseline covariates for possible inclusion in the primary analysis model where large imbalances are noted. However, due to the sample size, we do not anticipate substantial issues in this respect.

If sufficient data is available, for the PNC data secondary outcomes (Criminal offence data), we will adjust the corresponding secondary analysis model for a dummy indicator of school exclusion. This addresses secondary object point 6: Evaluate whether those excluded from school at the point of enrolment or during the trial will have a different rate of offending behaviour than those who have not been excluded.

6.2.3 Assumption checking

1. Linearity – plotting residuals vs predictor(s). If a structure is present, then transformation or an alternate model specification is required (i.e. GLM).

2. Homogeniety of variance – variance of the residuals across groups is the same. There is scope to fit models allowing for heterogeneous groups, but the setup is different (Generalized linear mixed model - GLMM).

3. Residuals are approximately normally distributed – plotting QQ plot

6.2.4 Alternative methods if distributional assumptions not met

If distributional assumptions are not satisfied, as appropriate, a generalized linear mixed model with alternate link function will be used.

The distributions of the primary outcomes will be assessed prior to conducting the analysis, if variables are skewed, then a Poisson mixed model will be specified, as follow:

$$g(Y_i) = \beta_0 + \beta_1 BL_i + \beta_2 TX_i + \beta_3 VIQ_i + \beta_4 Age_i + \beta_5 Sex_i + \beta_6 timewindow_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

Note: $g(.) = log_e(.)$, where $g(.)$ Is the log link function for the primary outcome measure.

Alternatively, data transformation could be used but use of the GLM is preferable.

6.2.5 Sensitivity analyses

Two types of sensitivity analysis will be conducted:

• Exploring the impact of missing data on trial outcomes by investigating likely missing data mechanisms and re-fitting the primary outcome within a multiple imputation framework (including exploring MAR and MNAR mechanisms via delta-based controlled multiple

imputation). Imputation variables for the model will include all covariates and the outcome appearing in the analysis as per recommendation by White, Woolstone & Wood (2011). In addition, variables that are predictive of missingness are included on the basis of strength of association with response variables. Also, any variables that explain response or non-response (Van Buuren, Boshuizen & Knook, 1999).

• Exploring the impact of different levels of intervention receipt on outcomes. We will use two-stage least squares instrumental variables (IV) regression to examine the effect of the intervention in those who receive varying levels of it. The proportion of sessions attended out of a maximum of six will be the instrumental variable in this analysis. The control group attendance will be set to zero and those intervention group will be assigned the number of sessions attended for the IV regression analysis.

Adherence will be categorised for the purposes of summary tabulation: attendance of ≥ 4 sessions (max number of sessions offered = 6). Following lower than expected attendance rates observed at 10/06/2025 TMG meeting, we have extended the CACE analysis to additionally include analyses with attendance defined at each level of at-least 1 through to 6 sessions.

Fidelity will be calculated by the average session score, then averaging across session to generate a single fidelity score. Fidelity items will be scored 0, 0.5, and 1. Total fidelity session score will be out of 18 or 20 depending on time point.

Both fidelity and adherence analyses will use a Two-Stage Least Square approach to estimate the model and Huber-White standard errors reported which are robust to clustering. The R packages 'ivpack' and 'ivreg' will be used to implement the two-stage instrumental variable analysis (Jiang & Small, 2014; Fox Kleiber, & Zeileis, 2021). Compliance (session adherence, i.e. number of sessions) will be instrumented by the intervention allocation (Angrist & Imbens, 1995). The stage 1 model is defined as follows:

$$Compliance_k = \beta_0 + \beta_1 TX_k + \varepsilon_{jk}$$

Predicted values for, $Compliance_k$, from the stage 1 model will be included in the stage 2 model, as follows:

$$Y_{ik} = \beta_0 + \beta_1 \widehat{compliance}_k + \beta_2 baseline_{ik} + \beta_3 Custody_k + \beta_3 VIQ_k + r_{ik}$$

6.2.6 Subgroup analyses

In addition to the primary and secondary outcomes, we have considered that the following outcomes may moderate the outcomes of this trial.

- Callous and Unemotional Traits: This will be measured, at baseline and 6 month follow up using the 24-item Inventory of Callous and Unemotional Traits – Parent/guardian Report and Youth Self-Report Versions (Essau et al., 2006) which are robust and well validated instruments (Ciucci et al., 2013). We will fit two moderation models for this variable to investigate the effect of moderation of treatment outcomes, but change may also occur as a consequence of treatment, so fitting both models permits us to disentangle these effects.

- Learning disabilities (LD): Children and young people will be invited to complete two subtests of the Wechsler Abbreviated Scale of Intelligence-II (WASI-II; Wechsler, 2011) to index their Verbal IQ. This scale is to be administered with a researcher (face-to-face, telephone, videoconferencing). The two subsets are to be included are Vocabulary (31 items) and Similarities (24 items). Raw scores are converted to scaled scores and summed, these are then age adjusted and a standardised score is created. The standardised score will be used in the moderation analysis.

- Retrospective or in-custody recruitment: This is a variable indicating whether the participant was recruited while in the custody suite or whether they were recruited retrospectively to the study within the 3 month window.

A moderation analysis will adjust the primary analysis with the inclusion of the moderator as a main effect and interaction between moderator and randomised group indicator. For example, the learning disabilities moderator analysis is as follows:

$$Y_i = \beta_0 + \beta_1 SDRM_{BL_i} + \beta_2 TX_i + \beta_3 VIQ_i + \beta_4 Age_i + \beta_5 Sex_i + \beta_6 timewindow_i + \beta_7 LD_i + \beta_8 TX_i * LD_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

where, Y_i are the SDRM scores; $SDRM_{BL}$ are the baseline SDRM scores; TX is the treatment/control variable indicator; *Custody* is the indicator of custody suite (strata, 8 levels: Blackburn, Preston, Blackpool, Harrow, Burnley, Hatfield, Stevenage and online).; *VIQ* is the Verbal IQ of the CYP (binary; ≤ 70 or >70); *Age* is the continuous measure of age in years, *Sex* is the biological sex indicator of the adolescent; *timewindow* is a binary indicator to distinguish if time to randomisation was “within time window” or “outside time window”; LD_i is learning disability status; $TX_i * LD_i$ is the interaction of learning disability status and treatment/control indicator; and ε_i is the individual level variation.

6.3 Missing data

Exploring the impact of missing data on trial outcomes by investigating likely missing data mechanisms and re-fitting the primary outcome within a multiple imputation framework (including exploring MAR and MNAR mechanisms via delta-based controlled multiple imputation).

We will summarise the extent of missing data in all outcomes and their respective control variables. A full multiple imputation strategy will be used if more than 5% of data in the primary model is missing. Alternatively, we will impute if more than 10% of data for a single variable is missing. We will use the multiple imputation by chained equations approach via the mice package in R (Van Buuren and Groothuis-Oudshoorn, 2011) and generate at least 10 imputed datasets, but will be proportionate to the percentage of missingness (i.e. larger proportions will have more imputed data sets generated). We will then estimate the intervention effect for each imputed dataset and pool the results using Rubin's combination rules for standard errors.

6.3.1 Missing data in item level data

The primary outcome measure's total score will be imputed directly. For secondary outcomes, we will not impute the PNC data as this will be assumed to be complete and counts of offences will not be imputed. Total scores of the SDQ will be imputed directly given that specific scoring rules for item level missingness are provided by the developers (see section 6.1.5). For the GARM measure, any missing item level data will be imputed using the chained equation approach, and imputed items summed for each imputed dataset to get total score per imputed dataset. Each item's imputation model will use other items and covariates specified in the analysis model as predictors.

Following creation of the imputed datasets, the corresponding total scores will be calculated using the imputed item level data. All imputed datasets will then fit the primary and secondary models and pool estimates following Rubin's rules.

Primary outcome

Given that each item is a count, we will use a Poisson regression (or negative binomial, if over dispersed) within the imputation model for each item.

Secondary outcome

Similarly, the correct link function will be used according to the item's structure for each of the secondary outcomes, i.e. binary or categorical accordingly. Therefore, a logistic or ordinal model will be used in the imputation for these items.

6.4 Additional analyses

Clustering via multilevel model

We will additionally consider the role of therapists as a source of clustering. As therapists will deliver the intervention to individuals allocated to the intervention arm only, this will be a form of partial nesting and may lead to an underestimation of standard errors (and thus inflated Type-I error) if not appropriately accounted for. We will also report intra-cluster correlation coefficients, the number of clusters, and cluster sizes. To account for any clustering, we will fit a heteroscedastic partially nested mixed-effects model structure (Candlish et al., 2018). The model will have a two-level structure, level 1 (individual) and level 2 (therapist). Verbal IQ, Age, Sex, and intervention variables will be included at level 1 and custody suite at level 2.

$$Y_{ij} = \beta_{0j} + \beta_{1j}SDRM_{BL_{ij}} + \beta_{2j}TX_{ij} + \beta_{3j}VIQ_{ij} + \beta_{4j}Age_{ij} + \beta_{5j}Sex_{ij} + \beta_{6j}timewindow_{ij} + U_jTX_{ij} + r_{ij}(1 - TX_{ij}) + \varepsilon_{ij}TX_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$U_j \sim N(0, \sigma^2{}_u)$$

$$r_{ij} \sim N(0, \sigma^2{}_r)$$

where, Y_{ij} are the SDRM scores; $SDRM_{BL}$ are the baseline SDRM scores; TX is the treatment/control variable indicator; *Custody* is the indicator of custody suite (strata, 8 levels: Blackburn, Preston, Blackpool, Harrow, Burnley, Hatfield, Stevenage and online. These will be included as fixed effects rather than a random intercept); *VIQ* is the Verbal IQ of the CYP (binary; ≤ 70 or >70); *Age* is the continuous measure of age in years, *Sex* is the biological sex indicator of the adolescent; *timewindow* is a binary indicator to distinguish if time to randomisation was "within time window" or "outside time window"; r_{ij} is the individual level variation in the non-clustered control arm; ε_{ij} is the individual level variation in the clustered arm; and U_j is the random intercept term for therapists.

In the first instance, we will assume compound symmetry as our correlation structure, but will investigate the autocorrelation plot and adjust the correlation structure as necessary, for example, first-order autoregressive (AR1) residuals.

Initially ICCs, at therapist level, will be calculated for the null model (without covariates predicting the SDRM); and then for the primary model (i.e. the model including the baseline SDRM score, Age, Sex, Verbal IQ, timewindow, and custody suite as covariates).

Longitudinal follow-up analyses

We will fit linear mixed models, accounting for repeated post-randomisation measures, SRDM outcome (6- and 12-months post-randomisation) within participants, adjusting for baseline measures, custody suite and counsellors to investigate the overall effect of the intervention on post-randomisation measures.

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}SDRM_{BL_{ijk}} + \beta_{2jk}TX_{ijk} + \beta_{3jk}VIQ_{ijk} + \beta_{4jk}Age_{ijk} + \beta_{5jk}Sex_{ijk} + \beta_{6jk}timewindow_{ijk} + U_{0k} + U_{1k}time_{ij} + \varepsilon_{ijk}$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$U_{0j} \sim N(0, \sigma^2{}_{u2})$$

$$U_{1j} \sim N(0, \sigma^2{}_{u1})$$

In this case, the effect size and 95% confidence interval will be calculated using given in Hedges (2007) for cluster randomised designed analysed via multilevel models and allowing for unequal cluster sizes. According to the two-level LMM for primary outcome, a sample estimate of the effect size equivalent to Hedges' g with 95% confidence interval is defined as:

$$\hat{\Delta}_g = \frac{\widehat{\beta_1}}{S_T}\sqrt{1 - \frac{2(n-1)\rho}{N-2}}$$

Where $\widehat{\beta_1}$ is the adjusted mean difference in SRDM score between trial arms; S_T is the within group pooled standard deviation (unconditional sample variance; 21)

$$S_T^2 = \frac{\sum_{i=1}^{m^I}\sum_{j=1}^{n_i^I}(Y_{ij}^I - Y_{i..}^I)^2 + \sum_{i=1}^{m^C}\sum_{j=1}^{n_i^C}(Y_{ij}^C - Y_{i..}^C)^2}{N-2}$$

Where 'm^I' is the total number of counsellors in the intervention sample, and 'n^I' the total number of participants (equivalent definitions apply for the control group, but with the 'C' designation). $Y_{i..}^I$ and $Y_{i..}^C$ are the mean outcomes among intervention and control counsellors respectively.

The remaining part of the $\hat{\Delta}_g$ equation makes the adjustment for clustering. The two intra - class correlation coefficients at the counsellor (ρ) level are defined as follows,

$$\rho = \frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2} = \frac{\sigma_B^2}{\sigma_T^2},$$

where σ_B^2 is the between-counsellor variance, and σ_W^2 is the within-counsellor variance. In addition.

For the remaining secondary outcomes, their effect sizes will be reported as either Hedges' g (Tobit, single level model and same as primary outcome) or rate ratios (all other secondary outcomes, exponentiated parameter estimates), given that generalized linear mixed effects models with log link function are used to model the data and that the measures are positively scored integers with some amount of skew anticipated (Barnett and Dobson, 2008).

6.5 Harms

The number (and percentage) of participants experiencing each AE/SAE will be presented for each trial arm categorised by severity. For each patient, only the maximum severity experienced of each type of AE will be displayed. The number (and percentage) of occurrences of each AE/SAE will also be presented for each trial arm. No formal statistical testing will be undertaken.

6.6 Statistical software

All statistical analyses will use R version 4.1.2 (2021-11-01) with additional packages: tidyverse, VGAM, lme4, lmerTest, performance, mice, psych, ivreg, and ivpack. Tobit models with random effects will be fitted using Stata 17 using the metobit function.

7. References

Joshua D. Angrist & Guido W. Imbens (1995) Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity, Journal of the American Statistical Association, 90:430, 431-442, 10.1080/01621459.1995.10476535

Barnett, A.G., & Dobson, A.J. (2008). An Introduction to Generalized Linear Models (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315182780

Bollen, K.A. and Stine, R. (1990). Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability. Sociological Methodology. Vol. 20:115-. DOI: 10.2307/271084

Borm G.F., Fransen J., Lemmens W.A. (2007). A simple sample size formula for analysis of covariance in randomized clinical trials. J Clin Epidemiol, 60(12):1234-8. doi: 10.1016/j.jclinepi.2007.02.006.

Candlish, J., Teare, M.D., Dimairo, M. et al. Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. BMC Med Res Methodol 18, 105 (2018). https://doi.org/10.1186/s12874-018-0559-x

Ciucci E., Baroncelli A., Franchi M., Golmaryami F.N., Frick P.J. (2013). The Association between Callous-Unemotional Traits and Behavioral and Academic Adjustment in Children: Further Validation of the Inventory of Callous-Unemotional Traits. Journal of Psychopathology and Behavioral Assessment, 36(2):189-200. Page 69 of 72 Version 1.3 25.10.2022

Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: a review of child self-report measures. Child and adolescent psychiatry and mental health, 8(1), 14.

Essau C.A., Sasagawa S., Frick P.J. (2006). Callous-unemotional traits in a community sample of adolescents. Assessment, 13(4):454-69.

John Fox, Christian Kleiber and Achim Zeileis (2021). ivreg: Instrumental-Variables Regression by '2SLS', '2SM', or '2SMM', with Diagnostics. R package version 0.6-1. https://CRAN.R-project.org/package=ivreg

Fonagy, P., Butler, S., Cottrell, D., Scott, S., Pilling, S., Eisler, I., Fuggle, P., Kraam, A., Byford, S., Wason, J., Smith, J. A., Anokhina, A., Ellison, R., Simes, E., Ganguli, P., Allison, E., & Goodyer, I. M. (2020). Multisystemic therapy versus management as usual in the treatment of adolescent antisocial behaviour (START): 5-year follow-up of a pragmatic, randomised, controlled, superiority trial. The lancet. Psychiatry, 7(5), 420–430. https://doi.org/10.1016/S2215-0366(20)30131-0

Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. Journal of Educational and Behavioral Statistics, 32(4), 341–370. https://doi.org/10.3102/1076998606298043

Jiang, Y. and Small, D. (2014). ivpack: Instrumental Variable Estimation. R package version 1.2. https://CRAN.R-project.org/package=ivpack

Otto, C., Kaman, A., Erhart, M., Barkmann, C., Klasen, F., Schlack, R., & Ravens-Sieberer, U. (2021). Risk and resource factors of antisocial behaviour in children and adolescents: results of the longitudinal BELLA study. Child and adolescent psychiatry and mental health, 15(1), 61. https://doi.org/10.1186/s13034-021-00412-3

Raby, C., & Jones, F. (2016). Identifying risks for male street gang affiliation: a systematic review and narrative synthesis. The Journal of Forensic Psychiatry & Psychology, 27(5), 601-644. doi:10.1080/14789949.2016.1195005

Raby, C., Jones, F.W., Hulbert S., & Stout, J. (2017). Design, development and validity testing of the gang affiliation risk measure (GARM), The Journal of Forensic Psychiatry & Psychology, 28:6, 753-776, DOI: 10.1080/14789949.2017.1324578

Senn, S. (1994), Testing for baseline balance in clinical trials. Statist. Med., 13: 1715-1726. https://doi.org/10.1002/sim.4780131703

Smith, D.J., & McVie, S. (2003). Theory and method in the Edinburgh study of youth transitions and crime. British Journal of Criminology, 43(1), 169-195.

Smith, D. J., Shute, J., Flint, J., McVie, S., Woodward, R., & McAra, L. (2001). *The Edinburgh Study of Youth Transitions and Crime: Key Findings at Ages 12 and 13*. Centre for Law and Society, University of Edinburgh.

Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. DOI: 10.18637/jss.v045.i03.

Tobin, J. (1958). Estimation of relationships for limited dependent variables, Econometrics, 26, 24-3

Twisk, J., Rijmen, F. (2009). Longitudinal tobit regression: A new approach to analyze outcome variables with floor or ceiling effects, Journal of Clinical Epidemiology, 62(9), 953- 958, https://doi.org/10.1016/j.jclinepi.2008.10.003.

Wechsler D. Wechsler Abbreviated Scale of Intelligence - Second Edition. Bloomington, United States: Pearson; 2011.

8. SAP Deviation Log

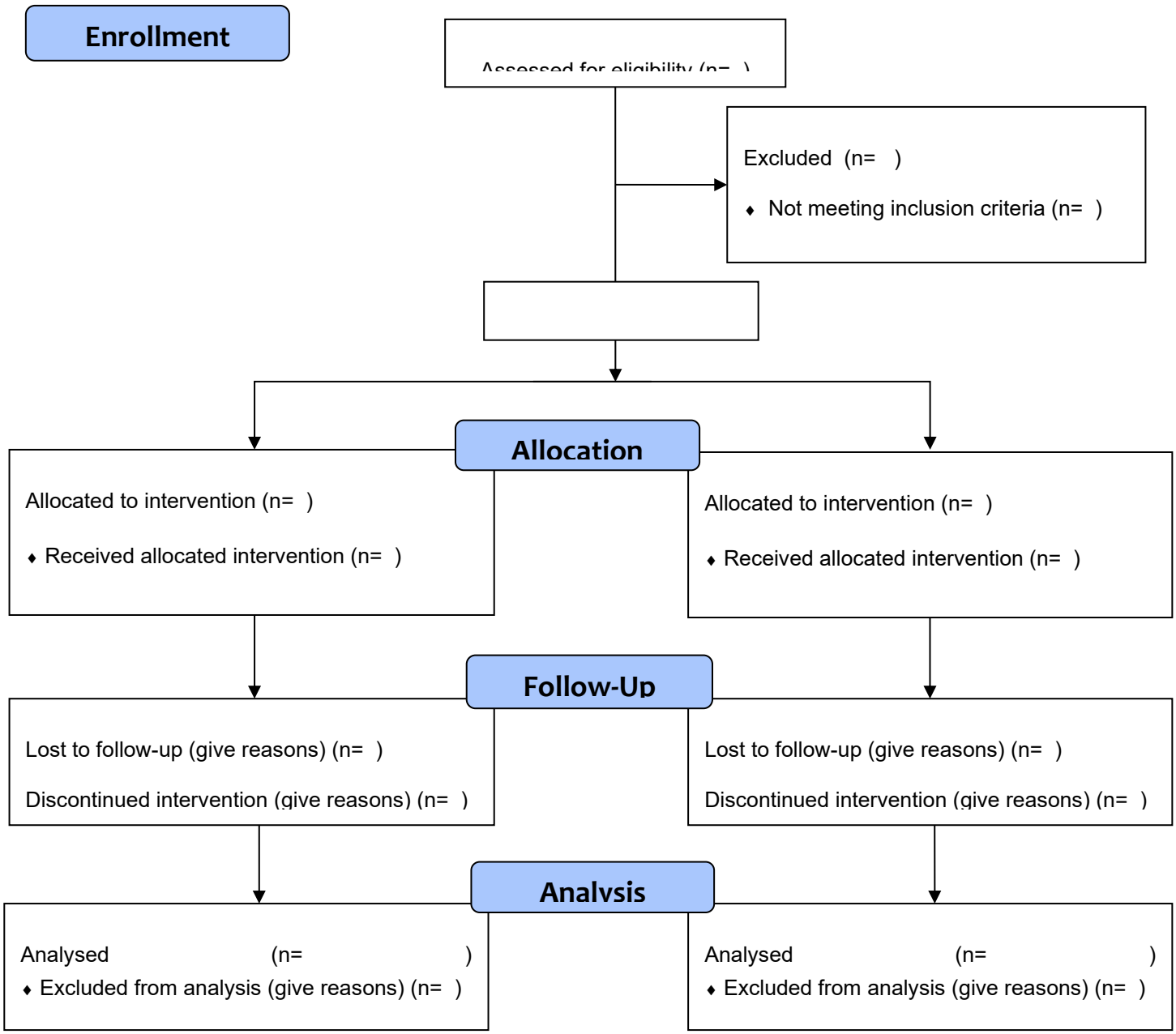
Document number:		**Document version:**	
Reason for deviation:			

Appendix A

CONSORT

TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram



Appendix B: Results tables templates

Table B1. Baseline characteristics of trial arms as randomised

Individual-level	Whole Group	Intervention group	Control group

(categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Sex (Female)							
Custody							
Blackburn							
Preston							
Blackpool							
Barrow							
Lancaster							
ONLINE							
Individual-level (continuous)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size (g)
Age							
VIQ							
Baseline SRDM							
Baseline GARM							
Baseline Arrests							
Baseline Cautions							
Baseline Reprimands							
Baseline Warnings							
Baseline Convictions							

Table B2. Baseline characteristics based on data completeness

Individual-level	Complete cases		Lost to follow up	
(categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Sex (Female)				
Custody				
Individual-level (continuous)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
Age				

VIQ				
Baseline SRDM				
Baseline GARM				
Baseline Arrests				
Baseline Cautions				
Baseline Reprimands				
Baseline Warnings				
Baseline Convictions				

Table B3. Primary analysis model coefficients

	Primary analysis	
Coefficient	*Estimates*	*CI*
Intercept		[,]
SRDM Baseline		[,]
Custody suite 2		[,]
Custody suite 3		[,]
VIQ		[,]
Age		[,]
Sex		[,]
Trial arm		[,]
Observations		
R^2 / adjusted R^2	__ / ___	

** p<0.05 ** p<0.01 *** p<0.001*

Table B4. Subgroup analysis model coefficients

	LD subgroup analysis		**Callous and unemotional traits subgroup analysis**	
Coefficient	*Estimates*	*CI*	*Estimates*	*CI*
Intercept		[,]		[,]
SRDM Baseline		[,]		[,]
Custody suite 2		[,]		[,]
Custody suite 3		[,]		[,]
VIQ		[,]		[,]
Age		[,]		[,]
Sex		[,]		[,]
Trial arm		[,]		[,]
LD x Trial Arm		[,]		
CALLOUS x Trial Arm				[,]
Observations				
R^2 / adjusted R^2	___ / ___			

** p<0.05 ** p<0.01 *** p<0.001*		

Table B5. Two level model for additional analysis allowing for clustering of therapists in intervention arm.

	Primary analysis(two-level)	
Coefficient	*Estimates*	*CI*
Intercept		
SRDM Baseline		
Custody suite 2		
Custody suite 2		
VIQ		
Age		
Sex		
Trial Arm		
Random Effects		
σ^2		
τ_{00}	___therapist_ID	

ICC	
N	__ therapist_ID
Observations	
Marginal R^2 / Conditional R^2	___ / ____
** p<0.05 ** p<0.01 *** p<0.001*	

Table B6. Primary analyses summary table.

Outcome Measures	Baseline				6 month follow-up				[a]Unadjusted mean difference	[b]Adjusted mean difference (covariates)	[c]Adjusted mean difference (therapist clustering)
	SAU		SAU + Intervention		SAU		SAU + Intervention				
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Difference (95% CI), p value	Difference (95% CI), p value	Difference (95% CI), p value
SRDM											
GARM											
Criminal Offences											
Arrests											
Cautions											
Reprimands											
Warnings											
Convictions											

[a]single-level model adjusted for baseline score, custody suite, and VIQ.
[b]Single level model adjusting for sex, age, VIQ, custody suite, and baseline score
[c]Multilevel model adjusted for therapist clustering (intervention arm only), VIQ, custody suite, and baseline score
** p<0.05 ** p<0.01 *** p<0.001*

Table B7. Missing data analysis using imputed data sets.

Outcome Measures	Baseline				6 month follow-up				[a]Unadjusted mean difference
	SAU		SAU + Intervention		SAU		SAU + Intervention		
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Difference (95% CI), p value
SRDM									
GARM									
Criminal Offences									
Arrests									
Cautions									
Reprimands									
Warnings									
Convictions									

[a]single-level model adjusted for baseline score, custody suite, and VIQ.
** p<0.05 ** p<0.01 *** p<0.001*

Table B8. Instrumental variable analysis for fidelity and adherence

Coefficient	Primary		IV analysis (fidelity)		IV analysis (Adherence)	
	Estimates	CI	Estimates	CI	Estimates	CI
Intercept		[,]		[,]		[,]
Baseline SRDM		[,]		[,]		[,]
Custody suite 2		[,]		[,]		[,]
Custody suite 3		[,]		[,]		[,]
Trial Arm		[,]		[,]		[,]
VIQ		[,]		[,]		[,]
Age		[,]		[,]		[,]
Sex		[,]		[,]		[,]
observations						
R^2 / Adjusted R^2						

** p<0.05 ** p<0.01 *** p<0.001*


CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
Centre for
Trials Research
Canolfan
Ymchwil Treialon


UKCRC
Registered
Clinical
Trials Units


WARWICK
THE UNIVERSITY OF WARWICK


NHS
Lancashire &
South Cumbria
NHS Foundation Trust

Appendix C: Questionnaires

See attached PDF files: "***Solutions - Child CRF V1.5 24.01.2023.PDF***" and "***Solutions - Parent CRF V1.4 24.01.2023.PDF***"

Appendix D: Progression criteria

Recruitment

(i) Up to 50% of overall target (n=222) (CYP) within first 7 months of recruitment to the trial (green=80 to 100%; amber=60 to 79%; red=<60%);

1400 children have been referred to Liaison and Diversion across LSCFT's x 3 custody suites between 1st April 2021 and 31st May 2022. Local data indicates numbers are increasing. Liaison and Diversion as a service, have high engagement rates with children and young people. Around 80% of children referred to the liaison and diversion service, accept the offer of an assessment.

Randomisation

(i) Number of CYP randomised (of CYP consented green=≥90%; amber=50-89%; red=<49%).

Retention

(i) Number of CYP (of randomised) not explicitly withdrawn from the trial (at 6-months: green=≥80%; amber=50-79%; red=<50%);

(ii) Are the approaches to maximise retention acceptable to participants in this trial? (assessed qualitatively through interviews with a small sample of CYP and parents/guardians).

Fidelity and adherence

(i) Fidelity assessed according to a fidelity checklist (developed in collaboration with the delivery team, prior to the internal pilot – see later in Process Evaluation) (green=≥80 of sessions meet criteria; amber=50-79%; red=<50%);

(ii) Adherence: session attendance (green=≥ 66.6% of scheduled sessions attended; amber= < 66.6% and ≥ 50; red=<50%) (an average of the number of sessions)

Outcomes

(i) Willingness of CYP to participate in trial processes (data completeness for 6-month Self Report Delinquency Measure: green=≥75%; amber=50-74%; red=<50%).

How does SFBT differ from SAU?

(i) SAU data from intervention and control groups in the internal pilot will be examined for any overlap with the content of the SFBT intervention (assessed by SAU questions in baseline questionnaire, and qualitatively through interviews with a small sample of CYP and parents/guardians), and to;

(ii) examine whether SAU is similar in the intervention and control groups, with the data

collated from services received (assessed by SAU questions in questionnaires, and qualitatively through interviews with CYP, parents/guardians, and practitioners)

Appendix E: Randomisation strategy document

Randomisation Strategy for *SOLUTIONS*: *Solution Focused Brief Therapy for young people in contact with the criminal justice system*

Eudract/ISRCTN No:	*[Number]*	Version Number:	3

Randomisation Strategy

Based on trial/study protocol version: **V1.6 12.06.2023**

Revision History		
Updated version no.	Description and reason for change	Date changed
2	Updated to reflect changes to the protocol, sample size, and addition of extra custody suites	11/10/2023

ROLES AND RESPONSIBILITIES

Trial Statistician: Dr Paul Thompson			
Role: Trial Statistician (Research Fellow in Applied Statistics, CEDAR, University of Warwick)			
Date:	11.10.23	Signature:	

Short title:	Randomisation Strategy for SOLUTIONS		
SOP number:	TPL/008/5	Version:	1.0

Senior Statistician: Dr Rebecca Playle			
Role: Deputy Director of Statistics, Centre for Trials Research, Cardiff University			
Date:	11.10.23	Signature:	
Chief Investigator(s): Dr Samantha Flynn / Professor Peter Langdon			
Role: Assistant Professor (Co-PI) /Professor (Co-PI), CEDAR, University of Warwick			
Date:		Signature(s):	
			
Other non-signatory contributor: [Name]			
Role:			

TABLE OF CONTENTS

1. STUDY DESIGN

A two-arm individually randomised RCT [1] of SFBT plus SAU vs. SAU alone, running over three years, involving CYP (age 10-17 years) who have presented at a one of five police custody suites in the NHS Trust region. There will be a 6- and 12-month follow-up period, and a process evaluation and internal pilot.

2. UNIT OF RANDOMISATION

Individuals will be randomised to either control or intervention study arm. It is unlikely given the recruitment strategy that individuals will come into contact with each other and will have different reasons for appearing at custody suite, so risk of contamination is minimised.

3. NUMBER OF GROUPS

Two groups are present in the study, Support as usual (SAU) plus Solution Focused Brief Therapy (SFBT) vs. SAU alone.

4. NUMBER TO RECRUITMENT

The brief indicated that between 350 and 450 CYP could be included in the study. Sample size estimates indicate 448 CYP will be recruited allowing for 20% drop out (N=359). We will use varying block size between 2 and 6. Following a period of slow recruitment, an adjustment of total sample size required and addition of two new custody suites has altered the total CYP required to n = 282 (225 after 20% dropout). The revised sample size uses a revised pre-post correlations, r=0.5 (originally, r=0.334).

5. RANDOMISATION RATIO.

The units recruited will be randomised to the intervention/treatment and control group with a fixed 1:1 ratio.

6. TYPE OF RANDOMISATION

CYP will be randomised on a 1:1 basis to either the intervention or comparator arm using stratified permuted blocks, stratified by verbal IQ and by custody suite [2]. The randomisation system will be embedded within the study database (REDCAP), and outcome assessors and trial statisticians responsible for analysing the data will remain blind to allocation.

6.1 BLOCK RANDOMISATION

Random permuted blocks with varying block size (sizes 2,4,6) are generated using Stata version 17.0 with the 'ralloc' function.

2. STRATIFICATION/BALANCING VARIABLES

Stratification variable(s) include custody suite (5 suites) and Verbal IQ (2 Verbal IQ levels: <=69 or >69) [2]. Noting that stratification is now across five sites not the original three sites. We have added a further ten randomization lists to account for potential future custody suites added if recruitment continues to be behind planned accrual.

6.3 NUMBER OF RANDOMISATION TO PREPARE

Originally, the randomisation script prepared 2 randomisation lists containing allocations required for the n=448 participants to be randomised (6 strata: 3 custody suite x 2 Verbal IQ levels: <=69 or >69). The first list was used to test the REDCAP system. The second list (which must contain different random allocations to the test list) is the list to be used for the trial.

Given the addition of two new custody suites and an adjustment to the overall sample size, a further list of allocations for the extra two suites will be appended to the existing allocations list. Also, planning for potential future custody suites, we have added a further ten lists (totaling twelve additional lists after inclusion of the two recruited custody suites). Appending to the original list ensures that existing allocations of recruited individuals are maintained. This additional list will contain 24 strata: 12 custody suite x 2 Verbal IQ levels: <=69 or >69, and will be appended to original list (totaling 30 strata: 15 custody suite x 2 Verbal IQ levels: <=69 or >69).

The REDCAP documentation recommends creating lists with the total number of allocations required per strata. The original listhad 6 strata (3 sites and 2 levels of IQ) with 448x6=2688 allocations in order to accommodate any eventuality of site or IQ imbalance under the original trial setup. The additional custody suite allocations will have 282x24=6768. When appended to the original list, the total allocations will total 9456 and will accommodate all possible combinations under the new recruitment conditions (addition of two custody suites).

6.4 SELECTION OF FINAL ALLOCATION

Not applicable - the randomisation list will be uploaded into the study database, and recruiters, outcome assessors and trial statisticians responsible for analysing the data will remain blind to allocation.

6.5 RELATIVE WEIGHTING FACTORS

No weights will be used, assuming equal weighting

7. ALLOCATION CONCEALMENT

A computer generated, random permuted blocked randomisation list is used to randomise participants. This list cannot be accessed by those recruiting participants and cannot be guessed or manipulated and so allocation concealment is maintained.

8. BLINDING

Trial statisticians responsible for analysing the data will remain blind to allocation. Participants and practitioners will not be masked to treatment arm. Our team will be masked to treatment arm; the exception will be the Trial Manager and researchers responsible for undertaking process evaluation interviews. Participants will be informed about the importance of maintaining masking. If inadvertent unmasking occurs, then this will be recorded and reported. We will work with the delivery team to develop a Standard Operating Procedure governing masking for those who are collecting data within the field which includes a protocol for dealing with inadvertent unmasking (e.g., reassigning research workers). Research staff will be blinded to allocation when collecting outcome measurements at both baseline and follow up; however, participants will not be blind to their allocation as this is not possible.

9. FALLBACK PROCEDURES IN CASE OF PRIMARY SYSTEM FAILURE

A copy of the allocation list used in the randomisation system will be kept securely, so that in the event of primary system failure the senior statistician or trial manager can manually allocate using the list. If access to the randomisation list is also lost, simple random allocation within sites irrespective of IQ will be used until the primary system is back up and running.

10. IMPLEMENTATION OF DESIGN

The randomisation lists (the test list and the trial list) will be created using STATA code written by the trial statistician. The test list will be used by the database designer and the trial statistician to ensure the allocation system is working. Once the randomisation system is tested and signed off as working correctly, the trial randomisation list will be uploaded to the study database. Cardiff CTR data manager will be responsible for maintaining the automated system and record of allocations. The record of allocations will be maintained in real time and provided to the trial team in the event of the primary system failure. The trial statistician will remain blind to the allocations until unblinding is required. This will occur after the study has closed for data collection and been analysed, or in the event, that the trial is stopped early.

11. TESTING THE RANDOMISATION ALGORITHMS AND SYSTEMS

The randomisation will be implemented and tested with simulated dataset at least five times and the results recorded in the relevant section of the TMF with the randomisation documentation. Further

checks to ensure treatment group numbers are evenly balanced at the end of each block will be conducted.

Testing stratification variables - The trial statistician will check that individuals are entering the correct strata in the database. Any discrepancies will be re-programmed and re-tested.

12. RISK OF SUBVERSION

The risk of subversion is minimal as the delivery team and individuals receiving SFBT are blinded to the allocations until the point of randomisation.

13. REFERENCES

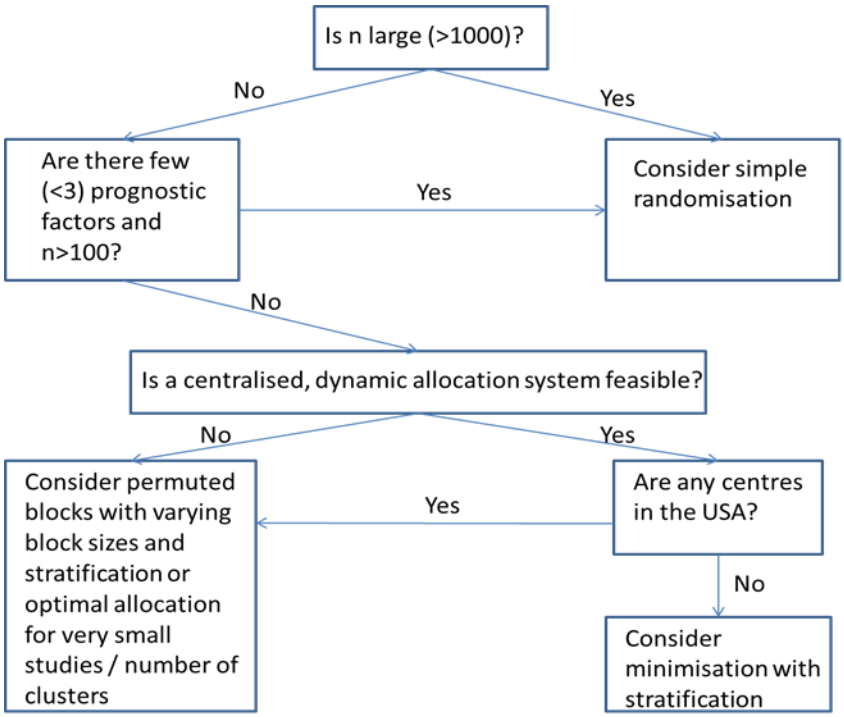
[1] ICH E9 Expert working group: ICH – Harmonised Tripartite guidelines: Statistical principles for clinical trials. Statistics in Medicine, 1999; 18: 1905-1942.

[2] Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trials. Biometrics 1975; 31:103-115

Appendix I - Additional guidance for choosing a method for randomisation

Checklist for choosing a randomisation strategy (adapted from [9])

- Is the trial individually or cluster randomised?
- How many subjects and study sites are planned?
- Are 24 hour randomisation services required?
- How will randomisation be implemented: central, remote local, face to face?
- Who will generate the sequence and by which method: random number lists, computer?
- Is a simple or stratified randomisation required?
- If stratified, how many strata and levels within each stratum are required?
- What balancing strategy should be chosen: simple, permuted blocks, minimisation, optimal balancing?
- What measures will be taken to guarantee allocation concealment?
- Who is going to monitor successful implementation (the balance of intervention group allocation, unblinding rates) during recruitment?
- Should a screening log of eligible subjects be collected to ensure participants are not excluded by foreknowledge of intervention group allocation?



Choosing a sequence generation method flowchart (adapted from [10,11])

The above flowchart does not cover all scenarios and is not prescriptive

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413