

STATISTICAL ANALYSIS PLAN

**Efficacy evaluation of the EXODUS
mentoring programme**

Coram

Principal investigator: Hannah Lawrence

Randomised controlled trial of the EXODUS mentoring programme

Statistical Analysis Plan

Evaluating institution: Coram

Principal investigator(s): Hannah Lawrence



Project title ¹	Efficacy evaluation of the EXODUS mentoring programme
Developer (Institution)	UpskillU Ltd
Evaluator (Institution)	Coram
Principal investigator(s)	Hannah Lawrence
SAP author(s)	Narendra Bhalla, Emily Blackshaw, Max Stanford, Hannah Lawrence, Mia Johnson Hall
Trial design	Two-armed randomised controlled trial (RCT), randomised at the individual level
Trial type	Efficacy trial with internal pilot and implementation and process evaluation (IPE) The internal pilot will begin in March 2024 with the end of the efficacy trial planned for the end of July 2027 and reporting in early 2028.
Evaluation setting	Community settings and secure training centres
Target group	11 to 17 year-olds (at time of referral) who are at least one of the following: • Known to be currently offending. Where CYP have been arrested or identified by police for offending

	and/or antisocial behaviour or affected by serious violence, criminal or sexual exploitation (as perpetrators and/or victims); or • Considered to demonstrate high/medium risk factors associated with offending, exploitation, and/or victimisation. This is based on vulnerability assessments undertaken by referral partners which considers whether the child or young person is at least one of the following: – Known to the Youth Offending Services (YOS) – Has a pending or in place National Referral Mechanism (NRM) as at risk of exploitation or trafficking – Has had a recent recorded missing episode – Known by the police/YOS to be affiliated with groups, often referred to as gangs, involved in crime, violence and trafficking – Known by the police/YOS to have siblings already involved in and affected by serious youth violence – Known to local authority children’s services (e.g. known to early help, is a Child in Need, on a Child Protection Plan, is Looked After, or has care experience) – Demonstrates exclusion risk factors i.e. persistent absences and suspension, displaying anti-social Behaviour Currently in or leaving custody. Where children and young people are in custody and will complete their custodial sentence within a given timeframe to allow them to engage in EXODUS.
Referral Partners	• Northamptonshire Councils, Police, Statutory Organisations and local schools

	• Haringey London Borough Council, Police, Statutory Organisations and local schools • Lewisham London Borough Council, Police, Statutory Organisations and local schools • Croydon London Borough Council, Police, Statutory Organisations and local schools • Oakhill Secure Training Centre
Number of participants	754 children and young people (CYP) recruited to the RCT
Primary outcome and data source	Self-reported offending behaviour 12 months post-date of randomisation Based on: Volume score on the Self-Reported Delinquency Scale (SRDS)
Secondary outcome and data source	1. Criminal offending 2. Violent offending 3. Criminal exploitation and victimisation 4. Emotional and behavioural problems 5. Mentor-mentee relationship 1. Local police data on the number of criminal offences 2. Local police data on the number of violent offences 3. Module A (Conventional Crime) of the Juvenile Victimization Questionnaire (JVQ). Secondary measure: local police data on number of incidents of victimisation and exploitation 4. Total difficulties score on Strengths and Difficulties Questionnaire (SDQ) 5. Social Support and Rejection Scale (SSRS)

SAP version history

Version	Date	Changes made and reason for revision
1.3	24.10.2025	Revisions based on adjustments made during extended pilot and transition phase
1.2	18.03.2025	Feedback from YEF external peer review
1.1	04.09.2024	Internal feedback from YEF
1.0 [original]	15.07.2024	

Table of contents

SAP version history	4
Table of contents	5
Introduction	6
Design overview.....	7
Sample size calculation	11
Clinically meaningful change	11
Pre-post-test correlation	11
Attrition and rolling recruitment	12
Planned sample size.....	12
Update to the sample size	13
Analysis	15
Internal pilot study analysis	15
Full efficacy study.....	21
Primary outcome analysis.....	22
Secondary outcome analysis.....	24
Subgroup analyses	27
Further analyses.....	28
Longitudinal follow-up analyses	29
Imbalance at baseline	30
Missing data	31
Compliance	32
Presentation of outcomes	34
References	36

Introduction

This document outlines the statistical analysis plan for the efficacy trial of EXODUS. EXODUS is a 12-month therapeutic restorative mentoring intervention where a young person aged between 11 and 17 is taken through a structured programme by paid mentors trained in restorative practice. It works with young people who are identified by statutory referral partners as demonstrating risk factors associated with violence and exploitation or as affiliated or affected by violence and exploitation. It aims to reduce the likelihood of children and young people (CYP) being affected by violence, offending and/or exploitation. For those leaving custody, it aims to support their reintegration into society. The programme is delivered by UpskillU, a fully restorative, community-based organisation and an established national restorative justice training provider.

According to the Youth Endowment Fund (YEF) Toolkit², restorative justice has a moderate impact on preventing crime and violence. The type of restorative approach used in the EXODUS mentoring programme has been found to be effective in family support service provision, where families felt taking a restorative approach was acceptable, engaging and helpful (Williams, 2019). Further, in the context of youth violence, mentoring has shown signs of promise and to be moderately effective (Gaffney, Jolliffe and White, 2022). This is particularly the case for interventions like EXODUS, where mentors have themselves been through the issues experienced by the CYP (Creaney, 2018). Evaluating EXODUS is an opportunity to build the evidence base on the effect of a restorative mentoring programme on young people displaying risk factors associated with offending.

EXODUS is being evaluated through an integrated efficacy study with an internal pilot and an implementation and process evaluation (IPE). An internal pilot is a phase in a trial after which progress is assessed against pre-specified criteria (Herbert et al, 2019), presenting the opportunity to stop the trial if these criteria are not met. Data collected during the internal pilot phase contributes towards the final analysis of a trial, increasing cost-effectiveness. Interventions which are well-developed and have previous evidence of promise, such as EXODUS, are considered suitable for this type of trial. Our full plans for the evaluation of the EXODUS programme can be found in our [evaluation protocol](#), published in July 2024.

² The Toolkit provides an overview of existing research on approaches to preventing serious youth violence, and can be found here: <https://youthendowmentfund.org.uk/toolkit/restorative-justice>.

Design overview

This is a multi-site trial where participants are being recruited across five³ referral sites. The trial uses a two-armed randomised controlled trial (RCT) design, where children and young people (CYP) aged between 11 and 17 are referred to EXODUS and are randomised to either the intervention or control group with a 50:50 allocation and on a rolling basis. The intervention group is referred to the EXODUS programme and the control group receive Business-As-Usual (BAU) support. Randomisation takes place at the individual level and is being stratified in five blocks (per referral partner) to ensure equal group allocation among participants from each referral site.

Participants are randomised at a baseline meeting with trained peer researcher, once they have provided written informed consent and completed the baseline outcome measures. Randomisation is completed by a Coram researcher using an easily operable Excel tool hosted securely on Coram's internal server and only accessible to the study team. The tool allows evaluators to see randomisation outcomes at the press of a button at the front end, while preventing them being able to anticipate the order of forthcoming randomisations hidden in the backend. Allocation concealment is being maintained, as those involved in enrolling participants (peer researchers, Coram researchers and referral partners) do not know in advance of randomisation how participants will be allocated.

Trial participants (CYP) are not blinded to which trial arm they are in, as they are aware from the consent procedures that they are taking part in a trial comparing business-as-usual support with a referral to EXODUS. The intervention delivery team (UpskillU) and referral partners delivering business-as-usual support are not blinded, as they know the CYP to whom they are delivering the EXODUS intervention. This lack of blinding is a necessity of the trial design. However, we have selected standardised CYP self-reported outcomes and routinely collected police data to avoid observer bias that may be introduced if outcomes were to be assessed by unblinded members of the delivery team or professionals from referral partners. Those undertaking the trial analysis will be blinded to randomisation. We will prepare the main analytical dataset so that trial arm is indicated by numbers and there is no data about participation in EXODUS (i.e. the data analyst cannot infer which participants received the intervention and which did not). This, in addition to the *a priori* data analysis plan, will prevent bias being introduced during data analysis.

³ The trial initially began with three referral partners, which expanded to four in April 2025 with the addition of the London Borough of Lewisham, and further to 5 in August 2025, with the addition of the London Borough of Croydon.

Table 1: Trial design overview

Trial design, including number of arms		Two-armed randomised controlled trial (RCT)
Unit of randomisation		Individuals (young people)
Stratification variables (if applicable)		Referral partner: • Northamptonshire Councils, Police, Statutory Organisations and local schools • Haringey London Borough Council, Police, Statutory Organisations and local schools • Lewisham London Borough Council, Police, Statutory Organisations and local schools • Croydon London Borough Council, Police, Statutory Organisations and local schools • Oakhill Secure Training Centre
Primary outcome	variable	Self-reported offending behaviour
	measure (instrument, scale, source)	Volume score on the Self-Reported Delinquency Scale (SRDS) (Smith et al., 2001) 12 months post randomisation
Secondary outcome(s)	variable(s)	1. Offending 2. Violent offending 3. Criminal exploitation and victimisation 4. Emotional and behavioural problems 5. Trusted adult relationship
	measure(s) (instrument, scale, source)	1. Local police data on the number of criminal offences, including the number of arrests (which lead to conviction), cautions, reprimands, warnings and convictions over 12 months post randomisation 2. Local police data on the number of violent offences over 12 months post randomisation 3. Module A (Conventional Crime) of the Juvenile Victimization Questionnaire (JVQ) (Hamby, 2004); Secondary measure: local police data on number of incidents of victimisation and exploitation

		4. Total difficulties score on the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) 5. Social Support and Rejection Scale (SSRS) (Roffman et al., 2000)
Baseline for primary outcome	variable	Self-reported offending behaviour
	measure (instrument, scale, source)	Volume score on the Self-Reported Delinquency Scale
Baseline for secondary outcome	variable	1. Offending 2. Violent offending 3. Criminal exploitation and victimisation 4. Emotional and behavioural problems 5. Trusted adult relationship
	measure (instrument, scale, source)	1. Local police data on the number of criminal offences, including number of arrests, cautions, reprimands, warnings and convictions in the 12 months prior to randomisation 2. Local police data on the number of violent offending in the 12 months prior to randomisation 3. Module A (Conventional Crime): Sub-scale of the Juvenile Victimization Questionnaire (JVQ) (Hamby, 2004) score prior to randomisation; Secondary measure: local police data on number of incidents of victimisation and exploitation 4. Total difficulties score on the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) 5. Social Support and Rejection Scale (SSRS) (Roffman et al., 2000)

As outlined in Table 1 above, the impact of the intervention will be assessed against the primary outcome of self-reported criminal offending after 12 months, with secondary outcomes of violent offending, criminal exploitation, emotional and behavioural problems, and mentor-mentee relationship also being assessed. Outcomes are assessed at 6, 12 and 18 (only in the case of local police data) months, but the primary findings will be based on the 12-month outcomes. Six-month outcomes are to inform decision making against the

progression criteria for the internal pilot study and 18-month outcomes are to exploratively assess longer-term efficacy.

Sample size calculation

Sample size predictions were calculated based on the primary outcome of volume of offending as measured by the Self-Report Delinquency Scale (SDRS) at 12-month follow-up. These figures were then reviewed and adjusted in discussion with YEF based on actual recruitment rates being observed in the extended transition phase within the pilot. However, we have first detailed the assumptions that supported our initial estimates below.

Clinically meaningful change

Previous research suggests that clinically meaningful change on the SDRS can be small, at just 4 (referenced in the evaluation protocol by Flynn et al., 2022) or 5 points (referenced in an evaluation of functional family therapy and UK youth offending; Humayun et al., 2017) of change on the SDRS. In development samples of the measure (N=4,106) the observed mean (M) was 8.3 with a standard deviation (SD) of 12.32 (Smith et al., 2001), indicating a MDES of 0.37 for 4.5-point change on the SDRS.

Unfortunately, previous research evaluating mentoring programmes and using the volume score on the SDRS is sparse. We identified one example in a small evaluation of the Youth at Risk's Coaching for Communities programme (Berry et al., 2009) where a mean difference of 5.8 points in the volume of offending was reported between intervention and control groups at follow-up. More broadly, a comprehensive meta-analysis of mentoring programmes for at-risk youth (Tolan et al., 2014) reported a standardised mean difference effect size of 0.21 (95% CI: 0.17-0.25) for delinquency (n=25 studies) measured in a variety of ways.

To detect change that is clinically meaningful and is in line with previously reported mentoring evaluations, we therefore based our initial sample size calculations on detecting effects of a small size (≤ 0.2) for the primary outcome.

Pre-post-test correlation

We included a baseline covariate adjustment in our power calculation as it is likely that for self-reported outcomes there will be a reasonable correlation. Unfortunately, there is very little reported on pre-posttest correlations for the SDRS in the existing literature. Although covering a broader topic area (health and related research at Sheffield university), a review of 20 RCTs using Patient Reported Outcome Measures (PROMS) reported that the majority of pre-post outcome correlations fell between 0.4 and 0.6 (Walters et al., 2019). A comparison of sample size scenarios varying this figure, with other assumptions held constant, is as below (Table 2). The more likely middling scenario of pre-posttest correlation being 0.5 was settled upon for the main sample size estimate, based on the cited literature.

Table 2: Sample Scenarios - Pre-Post Correlation

Scenarios (Varying pre-post correlation, other assumptions held constant: $ES = 0.2$, $Power = 0.8$, $alpha = 0.05$)	Sample Size (Power Calculations)	Sample Size Accounting for 30% Attrition)
Conservative (0.3)	718	1,026
Middling (0.5)	592	846
Optimistic (0.7)	402	574

Attrition and rolling recruitment

Data provided by UpskillU showed that dropout rates for an earlier 12-week iteration of the intervention were low (on average 5-7% dropout/non-completion for the three initial referral partners for referrals in 2022) and the aim would be to keep attrition below 10%, but we were cautious with these calculations, as these rates did not reflect drop-out for outcome measure completions. Based on our experience in previous randomised controlled trials (Taylor et al., 2023), we estimated dropout of outcome measure completions (for our primary outcome) as:

- 15% at six months and
- 30% at 12 months.

Planned sample size

As an initial estimate, we recommended recruiting a minimum sample of 846 (102 in pilot phase and 744 in efficacy phase) participants to detect an MDES of 0.2 in volume of offending at 12 months. The target sample size of 102 participants for the internal pilot was based on the timeframes guided by YEF coupled with UpskillU's projected recruitment rates.

Based on recruiting 20 or 34 CYP per month⁴ (deemed feasible by delivery partners), sufficient participants will have been recruited to detect an effect of this size by 31 July 2026.

⁴ From July 2024 to April 2025 the target drops to 20 per month.

Update to the sample size

Due to the challenges faced during the pilot, YEF granted an extension to the pilot until 1 July 2025. This decision was made in April 2025, to test whether mitigations implemented to address the issues around recruitment, retention, and quality of the monitoring data resulted in improvements in these areas. Based on this revision, the number of CYP in the pilot phase sample increased from 102 CYP to 164 CYP (recruited up to 31 December 2024).

During the extension period, we reviewed the expected sample size for the efficacy trial based on actual recruitment rates. At the time, 259 CYP had been randomised against a target of 366 CYP. Through discussion with UpskillU and YEF, we revised the expected estimates based on the mitigations put in place (e.g., addition of school-based referrals and Lewisham Council as a referral partner).

In the time and resources available for the efficacy study, we estimated that a minimum total sample size of 528⁵ would be more realistic, amounting to 754 CYP factoring in 30% attrition. (164 in pilot phase and 590 in efficacy phase). This amounted to recruiting 495 CYP or 33 CYP per month from the point of re-calculation (01 May 2025) until the end of the recruitment period (31 July 2026), a comparable number to the figures set out in the previous estimate. This sample would achieve an MDES of 0.212. However, we are working with UpskillU to make every effort to achieve a sample size as close as possible to 592 as this would be more desirable and in line with YEF guidance. Given the primary outcome is measured at 12 months post randomisation, the outcome measures data will be available by August 2027 (as per our assumption of 30% attrition).

Sample size calculations were carried out in R using the 'PowerUp' package, using the model with Individual-level random assignment design with fixed block effects. The initial and revised effect sizes and sample estimates are shown below in Table 3.

Table 3: Sample Size Calculations

		Protocol Parameters	Post-Pilot Parameters
Minimum Detectable Effect Size (MDES)		0.2	0.212
Pre-test/ post-test correlations	level 1 (participant)	0.5	0.5

⁵This target was still kept comparable to the original planned sample given that the referral trend saw a consistent increase from March to June 2025 with the additional sources starting to refer to the trial. A total of 157 CYP were randomised in this four-month period, as compared to 196 over the previous 11 months.

		Protocol Parameters	Post-Pilot Parameters
Intracluster correlations (ICCs)	Level 2 (referral partner)	-	-
Alpha ⁶		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size	Individuals within referral partners	197.3	105.6
	Referral partner count	3	5
Number of participants	Intervention	296	264
	Control	296	264
	Total	592	528

Analysis

EXODUS will be evaluated through a two-armed randomised controlled trial (RCT) efficacy study with internal pilot. Below we set out analysis planned for the internal pilot and then for the efficacy study.

Internal pilot

Study criteria and reporting

The internal pilot was carried out to establish the readiness of the EXODUS programme for a full efficacy trial. The primary research question was: “Is a full efficacy trial of the EXODUS intervention feasible?”

The internal pilot initially had a target of recruiting 102 participants (CYP), with the aim of recruiting 34 CYP per month during the first three months of the trial, from March 2024 to July 2024. Given the extension granted by YEF up to 1 July 2025, the extended pilot phase now covers 164 CYP, recruited from March 2024 to December 2024. As a result of extension, the analysis and reporting period for the internal pilot began in September 2025, with final publication due in March 2026.

The pilot aimed to assess:

- **Recruitment:** including the extent to which the referral pathways are working. We considered whether sufficient referrals are flowing into EXODUS, and whether these referrals are meeting eligibility requirements assessed via referral forms.
- **Randomisation:** the randomisation process including the acceptability of the referral pathways and consent and randomisation procedures to participants assessed via referral and drop-out rates.
- **Data collection:** processes for collecting data on the demographic profiles and characteristics of CYP randomised. Along with this, monitoring data on the uptake and attendance of those in the intervention group and what business-as-usual support looks like in the control group.
- **Response rates:** the feasibility of collecting outcome measures data via an online survey administered to CYP and local police data on (re)offending and violent (re)offending.
- **Fidelity:** measured against suggested fidelity sub-criteria outlined in Table 4 within criteria number 11 (‘11. Fidelity to programme: Deviations from the intervention logic model by delivery team’) and may be developed further during the internal pilot (see Compliance section for more detail). These sub-criteria reflect the core components

of EXODUS and have been further co-developed with UpskillU, evaluators and peer researchers.

- **Scale-up:** the ability of the EXODUS programme to scale-up for the full trial.

In line with standard practice for deciding on progression from an internal pilot to a full efficacy trial (Avery et al., 2017), we developed 12 criteria (Table 4) which set out the targets for stopping the trial ('stop'), reviewing the trial ('review') and continuing ('proceed') the internal pilot to an efficacy trial. The criteria were produced with input from UpskillU and delivery partners and are based on standard criteria used in previous internal pilots. We reviewed data collected in the internal pilot against the 'stop,' 'review,' and 'proceed' criteria in consultation with the trial steering group and our peer researcher young advisors.

As a guide:

- If at least ten out of the 12 the 'proceed' criteria are met, we would recommend proceeding with to the efficacy trial
- If there are at least ten 'review' criteria met, we would recommend reviewing the trial before proceeding. In this instance, we will recommend that the trial is paused for the time needed to make amendments until it can begin again
- If there are more than six 'stop' criteria met, we would not proceed to the efficacy trial. In this instance we will take a phased approach to ending the trial which will aim to avoid any negative impact on CYP.

Table 4: Internal Pilot progression criteria

Criteria	Green (proceed)	Amber (review)	Red (stop)
1. Recruitment rate: recruitment of participants (CYP) to internal pilot trial on track against target (n= 102)	80-100% (n=82-102) of target number of CYP participants recruited to internal pilot	60-79% (n=61-81) of target number of participants recruited to internal pilot	0-59% (n=0-60) of target number of participants recruited to internal pilot
2. Randomisation proportions: the proportions of participants allocated to each group should be 50:50	CYP are randomised into the control or intervention group 50:50	CYP are randomised into the control or intervention group 45:55	CYP are randomised into the control or intervention group 40:60

3. Eligibility rate: CYP referred to the trial are eligible	90-100% of CYP referred are eligible (as per eligibility criteria here)	70-89% of CYP referred are eligible	0-69% of CYP referred are eligible
4. Consent rate: CYP who meet the eligibility criteria give consent to take part in the trial	90-100% of eligible CYP give consent	70-89% of eligible CYP give consent	0-69% of eligible CYP give consent
5. Randomisation rate: CYP who consented to taking part in the trial have been randomised	90-100% of CYP who consented have been randomised	70-89% of CYP who consented have been randomised	0-69% of CYP who consented have been randomised
6. Response rate: response rate for primary outcome measure at six months for CYP in intervention and control group ⁷	85-100% completion rate	60-84% completion rate	0-59% completion rate
7. Attrition rate to the trial: low attrition rates for both arms of the trial (intervention and control group)	0-20% attrition rate on both arms of trial	21-35% attrition rate on both arms of the trial	36-100% attrition rate on both arms of the trial
8. Referral partner data: sufficient data collected from at least one referral partner on official police (re)offending and violent (re)offending	Low rates of missing data (80-100% of cases complete) from at least one referral partner	Some missing data (60-79% of cases complete) from at least one referral partner	High rates of missing data (0-59% of cases complete) from at least one referral partner

⁷ We recognise there will likely be higher response rates for the intervention group, but for both we expect at least an 85% completion rate.

9. Delivery partner data: sufficient monitoring data collected from UpskillU to allow for analysis of uptake and demographic profile of CYP	Low rates of missing data (90-100% of cases complete)	Some missing data (70-89% of cases complete)	High rates of missing data (0-69% of cases complete)
10. Fidelity to trial: deviations from trial protocol from referral partners and delivery team	Low rates of deviations from the protocol (80-100% of cases with no deviations) Measured by: Incidents of contaminations e.g. monitoring business-as-usual support to check for any reports of CYP accessing 12 month EXODUS	Some rates of deviations from the protocol (70-79% of cases with no deviations) Measured by: Incidents of contaminations e.g. monitoring business-as-usual support to check for any reports of CYP accessing 12 month EXODUS	High rates of deviations from the protocol (0-69% of cases with no deviations) Measured by: Incidents of contaminations e.g. monitoring business-as-usual support to check for any reports of CYP accessing 12 month EXODUS
11. Fidelity to programme: deviations from the intervention logic model by delivery team	Low rates of deviations from the logic model 80-100% of cases with no deviations Measured by monitoring UpskillU mentoring debrief and missed session debrief forms and using a fidelity checklist (also here) which includes: 1. Dosage: a) at least 60% of organised in-person and virtual mentoring sessions are attended by CYP (yes/no)	Some rates of deviations from the logic model 70-79% of cases with no deviations (measured in the same way as Green)	High rates of deviations from the logic model 0-69% of cases with no deviations (measured in the same way as Green)

	b) at least 50% of the planned COSA meetings are attended by CYP (yes/no) 2. Coverage and consistency: a) whether four or more core themes⁸ are covered in the 12-session intensive phase by mentor (yes/no) b) risk assessments are completed for at least 80% of sessions by mentor (yes/no).		
12. Acceptability: acceptability of trial design	There is a low level of reports from referral partners/other professionals of CYP requiring additional emotional support (e.g. from their YOS worker) after being randomised (for instance, if distressed about the randomisation outcome). This will be gathered from the de-brief session with peer researchers/UpskillU peer researcher manager (0-20% CYP required additional support from a referral partner professional)	There is a medium level of reports of CYP requiring additional support (e.g. from their YOS worker) after being randomised which will be gathered from the de-brief session with peer researchers (21 - 35% CYP required additional support)	There is a high level of reports of CYP requiring additional support (e.g. from their YOS worker) after being randomised which will be gathered from the de-brief session with peer researchers (36-100% CYP required additional support)

We will report on the trial's performance against these progression criteria in the pilot report. Additionally, we will report on progress and adjustments made in the pilot extension period and the performance against the additional set of criteria for that period detailed below.

⁸ These are: 1. Self-identity, 2. Attitudes and values, 3. Risks and consequences, 4. Conflict and choices, 5. Relationships, influences and decisions, 6. Solutions and strategies.

Additional progression criteria for the pilot extension period:

- 1. Improvement in retention following implementation of mitigation strategies. As of 1st of July:**
 - Observing close to 70% response rate on 6-month follow-up surveys due within the 6-week response window.
 - A minimum of 70% of CYP referred starting 1 March, including new referral pathways, attend mentoring sessions (treatment group) or respond to monthly check-ins (control group).
- 2. Improvement in recruitment rate and engagement following onboarding of additional referral partners. As of 1st of July:**
 - Lewisham council and school referrals lead to an increase in randomisations per month during April-July 2025.
 - The overall recruitment is on track to reach the minimum required sample size of 754 CYP randomised.
 - Suitability of referrals from the new partners is high, translating into high engagement.
- 3. Improvement in consistency and quality of the monitoring data following adoptions of the new Case Management System and addition of a data administrator. As of 1st July 2025:**
 - UpskillU independently keep up-to-date monitoring data on their CMS, and are able to successfully transfer monitoring and demographic data to the evaluator for 100% of the sample every 2 weeks, without delays.
 - Low rates of missing data in mentor logs (90%-100% of data is complete).
 - Coram is able to fully assess fidelity based on the monitoring data shared by UpSkillU.

Analysis

We will use descriptive statistics (means, standard deviations, percentages etc.) to understand trial recruitment rates, intervention delivery rates, and measure completion rates at baseline and 6-month post randomisation. We will also use descriptive statistics to analyse demographic characteristics of the sample in the internal pilot study to better understand who is engaging in the trial and identify missing data in routine data capture systems. Based on the referral form used by UpskillU, we anticipate analysis of the data on sex, age, and ethnicity; Special Education Needs and Disabilities (SEND); English is an additional language; care experienced or open to children's social care/early help; currently experiencing mental health problems; open to Child and Adolescent Mental Health Services (CAMHS) or Child Criminal Exploitation (CCE)/Child Sexual Exploitation (CSE) intervention; open to the Youth Justice Service; eligible for Free School Meals; and migrant/refugee status.

We will look at age using means, standard deviations, and ranges, and other categorical data using frequencies. We will look at this data in terms of those who consent and are randomised, those that complete baseline measures, and those with follow-up data available at 6-months. While we will review the balance of characteristics across trial arms, given the small numbers recruited at this early stage (164 CYP), no inferential statistics in terms of impact will be conducted at this stage.

Full efficacy study

If the internal pilot criteria are met, the study will progress to a full efficacy trial.

The primary research question for the efficacy trial is:

1. What is the difference in self-reported offending rates of CYP demonstrating risk factors or affected by offending or exploitation, between those who receive a targeted restorative mentoring programme and those who receive business-as-usual support in youth offending, custody and community safety services?

The secondary research questions are:

2. What is the impact of referring a CYP, affected or demonstrating risk factors associated with offending or exploitation, to a targeted restorative mentoring programme, relative to those who receive business-as-usual support, on:
 - Violent offending
 - Criminal exploitation and victimisation
 - Emotional and behavioural problems
3. Given the aim of the EXODUS programme to support the reintegration of CYP who are leaving custody into society, is the impact different for CYP who are leaving custody?
4. With CYP from marginalised and disadvantaged communities overrepresented in the programme, is the impact different for CYP within these groups having specific characteristics, particularly those from more marginalised⁹ groups?
5. Do EXODUS participants build high-quality relationships with their mentors? What is the relationship between high-quality mentor mentee relationships and offending rates, criminal victimisation, and emotional and behavioural problems? Is the impact on mentor-mentee relationships for those receiving EXODUS different relative to trusted adult relationships for those receiving business as usual support?

Analysis for the full efficacy study will include all randomised participants who provide outcome data across both the pilot and the efficacy study.

⁹ This could include girls and young women, CYP with mental health issues, migrant and refugee CYP, CYP from ethnic minority groups, care-experienced CYP, those with SEND and CYP who are economically disadvantaged.

All outcomes will be analysed on an intention-to-treat basis in that all participants will be analysed according to the trial arm to which they were assigned, as opposed to whether the intervention was received. We will report a baseline description of the trial participants using demographic data from UpskillU referral forms, referral partner monitoring data, and baseline outcome data. We will carry out balance checks to report on how the characteristics of respondents are balanced across treatment and control groups. These characteristics include those collected through the EXODUS referral form such as sex; age; ethnicity; SEND; English is an additional language; care experienced or open to children's social care/early help; currently having mental health problems; open to CAMHS or CSE/CCE intervention; open to the Youth Justice Service, eligible for Free School Meals; and migrant/refugee status.

We will report full baseline characteristics of the sample including baseline outcome scores, the characteristics of those lost to follow-up, and the characteristics of the analysable sample. We will also report participant flow throughout the trial, including completion rates of outcomes in a CONSORT diagram. Reasons for drop-out/missing data will be detailed where known.

For all analysis we will adhere to our good spreadsheet design principles which are tried and tested and will document the sequence of steps used to get from raw data to findings to enable review. All data cleaning and analysis will be undertaken in R statistical software. All code and analysis will be quality assured by a second member of the Coram study team, covering both the logic and the arithmetic of analysis. Full records of code will be shared with YEF and published to enable replication. Any deviations from the present SAP will be submitted as substantial amendments where applicable and recorded in subsequent versions of the protocol and SAP. In the final report, deviations and ad hoc/exploratory analysis will be highlighted for transparency.

Primary outcome analysis

The primary outcome of self-reported offending behaviour at 12-month follow-up is being measured using the Self-Report Delinquency Scale (SRDS; Smith et al., 2001), which assesses the frequency and severity of 19 offending behaviours¹⁰ in the previous 12 months. There is evidence that respondents answer accurately when asked if they have carried out these offending behaviours (Nock et al., 2006; 2007). Internal consistency of the measure has been reported as 0.87-0.92 and inter-item correlation has been reported as 0.19 (Fonagy et al., 2018; Humayun et al., 2017). The measure has been found to correlate with official police records of arrests (89.5%-95.2%) (McAra & McVie, 2005).

¹⁰ 18 of the 19 offences have a direct link to the established scoring structure for the questionnaire. However, alternative approaches to scoring will be considered with input from YEF and UpskillU for the final report.

The SRDS is being used as the primary measure of offending as it captures a substantial number of different types of offending behaviour. It is also more sensitive to changes in offending in the sample size anticipated for this trial compared with local police data. This is because local police data only captures instances reported by the police and some offending behaviour goes unrecorded, especially for CYP. The SRDS produces two scores: the volume of delinquency score (the estimated minimum total number of offending behaviours reported) and the variety of delinquency score (the number of different offending behaviours reported). We will be using the volume of delinquency score (a count/frequency variable) as our primary outcome owing to its greater sensitivity to change but will also analyse and report on the variety of delinquency score.

We will use an ANCOVA model to estimate the average effect of treatment allocation on offending. We will include fixed effects in the regression model to account for referral partner (stratifying variable). We will also adjust for baseline SRDS scores, sex and ethnicity.

The regression coefficient will provide an estimate of the size and direction of the treatment effect and its significance will be tested with a two-tailed 5% Type I error threshold. We will report our effect sizes as Hedge's *g* using unconditional (unadjusted) standard deviations accompanied by bootstrapped confidence intervals as a measure of uncertainty. The model is summarised as follows:

$$Y_i = \beta_0 + \beta_1(Arm_i) + \beta_2(Site_i) + \beta_3(Baseline_i) + \beta_4(Sex_i) + \beta_5(Ethnicity_i) + \varepsilon_i$$

Where, for the i^{th} young person:

- Where Y_i is the 12-month SRDS volume score;
- β_0 is the grand intercept and β_k , where $k = \text{no. of predictors}$, are regression coefficients;
- Arm_i is a binary variable representing trial arm allocation;
- $Site_i$ represents the vector for the referral site dummy variables (for the 5 sites, 4 binary dummy variables will be included);
- $Baseline_i$ is the baseline SRDS volume score;
- Sex_i and $Ethnicity_i$ of the CYP have also been included in the model as covariates; and
- ε_i is the error term.

We anticipate that the primary variable of volume of offending may be positively skewed owing to the count nature of the data. We will assess skew by summarising the distribution of the data using plots and skew statistics, as well as checking model residuals. We will make a final decision about the best approach to address skew once this has been assessed, but this may include transforming the data with a log link function (Poisson or negative binomial).

We will check for balance across the demographic variables listed in Table 6 within the 'Imbalance at baseline section'.

Secondary outcome analysis

In line with the aims of EXODUS and its theory of change, the secondary outcomes being analysed include reducing recorded incidence of offending and violent offending, as well as self-reported criminal exploitation and emotional and behavioural problems. We will also analyse the quality mentor-mentee relationship given it is a key aim of the EXODUS programme. These measures are described in more detail below and are set out in Table 5.

Table 5: Secondary outcome measures for the EXODUS evaluation

Secondary outcomes	Measure	Scoring Range	Source	Data Collection Timeline
Recorded incidence of offending	Local police data, including arrests, cautions, reprimands, warnings, and convictions.	-	Administrative Data from referral partners	Baseline, 6 months, 12 months, 18 months (data for the previous 12 months)
Recorded incidence of violent offending	Local police data, including arrests, cautions, reprimands, warnings, and convictions associated with violent offending	-	Administrative Data from referral partners	Baseline, 6 months, 12 months, 18 months (data for the previous 12 months)
Criminal victimisation and exploitation	Sub-scale of the Juvenile Victimization Questionnaire (JVQ) (Module A: Conventional Crime) Secondary measure: Local police data on number of incidents of victimisation and exploitation.	0 to 8	Respondent survey	Baseline, 6 months, 12 months
Emotional and behavioural problems	Strengths and Difficulties Questionnaire (SDQ)	0 to 40	Respondent survey	Baseline, 6 months, 12 months
Trusted adult relationship	Social Support and Rejection Scale (SSRS) with four sub-scales:	S1, S4: 1 to 6	Respondent survey	6 months, 12 months

	- Feeling Valued (S1): 1-6 - Trust (S2): 7-11 - Mentoring (S3): 12-16 - Negativity (S4): 17-22	S2, S3: 1 to 5 (average across items in each sub-scale)		
--	---	--	--	--

Local police data

While the SRDS is quite sensitive to minor offences that might not show up on a police database, due to the self-report nature of the questionnaire, the CYP may not report all offending behaviour through the SRDS due to error or unwillingness. Therefore, local police data is being used as a supplementary, secondary-outcome measure of **recorded incidence of offending**. This includes, for example, the number and type of criminal occurrences and their associated outcomes (whether the young person is still a suspect or has been convicted). Analysis will primarily focus on the volume of offending, but will also look at the type of offending. Data about CYP will be captured at baseline, 6 and 12 months and at 18 months for those eligible within the sample. 18-month data capture will allow for explorative assessment of longer term impacts of the intervention. It was included in the study after discussion with YEF, primarily based on the rich data easily available with at least one referral partner, which does not require any further burden on participants (for more on the analysis see section on longitudinal analysis). Data collected at each of these stages will look at offending data for the previous 12 months.

Given a focus of EXODUS and the YEF is CYP demonstrating risk factors associated with youth violence, a secondary outcome will be **recorded incidence of violent offending**. This is defined as any offence involving violence or threats of violence, ranging from assault to murder (Crown Prosecution Service, 2022). This is also being measured using local police data but focuses on the number and type of criminal occurrences under violent offending and their associated outcomes (whether the young person is still a suspect or has been convicted). Analysis will primarily focus on the volume of violent offending but will also look at the type of violent offending. Data will be captured a baseline, six and 12 months and at 18 months (as above) for those eligible within the sample (for more on the analysis see section on longitudinal analysis). At each stage, data will look at offending data for the previous 12 months.

Self-reported data

In line with EXODUS's intended outcome of reducing CYP's likelihood of being criminally exploited, **Criminal exploitation** is being measured directly from CYP using one sub-scale of

the **Juvenile Victimization Questionnaire (JVQ): Module A: Conventional Crime** (8 items). Guidance on the JVQ indicates that modules of the measure can be used individually where a more focussed assessment is required (Finkelhor et al., 2005). However, we have found limited reported psychometric properties for Module A: Conventional Crime when used independently (although the module has been reported to have α above 0.6 in Finkelhor et al., 2005). We will therefore report on internal consistency reliability (α) in our outputs. As we are using the 8 items in Module A, the maximum score will be 8. We will also report on the percentage of CYP who answer 'yes' to at least one item. The measure will be collected directly from CYP at baseline, 6-months, and 12-month follow-up.

In line with EXDOUS's intended outcome of reducing CYP behavioural problems and improving social emotional regulation, the trial is measuring **Emotional and behavioural problems** using the self-reported **Strengths and Difficulties Questionnaire (SDQ)**. The SDQ is being used by YEF across its projects to create consistency and comparability between different evaluations. The SDQ will be collected directly from CYP at baseline, 6-months, and 12-month follow-up. It is a brief behavioural screening questionnaire with 25 items across 5 sub-scales covering emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour. Our main analysis will focus on the total difficulties score (first 4 sub-scales listed above), but we will look at individual sub-scales as part of our exploratory analysis. The total difficulties score ranges from 0 to 40. The SDQ shows internal consistency (alpha coefficients = 0.79 to 0.80) (Haywood et al., 2014). The self-reported total SDQ score has a test-retest reliability score of 0.79 and a 4-6-month stability score of 0.62 (Achenback et al., 2008).

The CYP's **trusted adult relationship** is being measured using the 22 item **Social Support and Rejection Scale (SSRS)**. This scale is designed for a CYP (aged 10-18) to self-report their positive and negative interactions with significant non-parental adults. It assesses four dimensions of social support and social rejection that youth may experience in relationships with important non-parental adults. These dimensions are:

- **feels valued** (6 items, e.g., "This person cares about me even when I make mistakes.")
- **trust** (5 items, e.g., "I talk to this person about problems with my friends.")
- **mentoring** (6 items, e.g., "I learn how to do things by watching and listening to this person.")
- **negativity** (6 items, e.g., "I feel that this person will let me down.")

Response options are Never, Rarely, Sometimes, Often, or Always. It is being administered to CYP in the intervention and control groups at the 6- and 12 month follow-up (it will not be collected at baseline as the intervention group will not yet have access to a mentor). This aligns with the logic model that a trusted and meaningful mentor relationship can support CYP to make positive changes in their lives. We decided to use this measure based on the face

validity of the items and their relevance to the EXODUS mechanisms of change described by UpskillU. Given limited available evidence on test reliability, we will also report on internal consistency (α) in our outputs. While on the intervention side the CYP is likely to answer the questions with reference to the EXODUS mentoring relationship, the questions are phrased more generally across trial arms to keep things consistent for comparisons.

The SSRS will be collected directly from CYP at baseline, 6-month, and 12-month follow-up. Each item is scored from 1 (Never) to 5 (Always) and each subscale score is the average of the items that make up the subscale. Higher scores on the 3 positive scales reflect higher levels of support within the relationship, while higher scores on the negativity scale reflect higher levels of stress and negativity within the relationship. Given the disaggregated interpretations, all four sub-scales will be scored and analysed separately, as combining them may lead to a loss of information that each scale will provide.

Analysis Approach

All models (as well as the SRDS variety of offending score) will be analysed according to the approach specified for the primary outcome, i.e. we will use an ANCOVA regression model¹¹. All models will include fixed effects to account for referral partner (stratifying variable). We will also adjust for baseline values of the analysed outcomes, as well as sex and ethnicity. In the analysis of the JVQ, to address possible ceiling/floor effects that may arise, we will adjust to a censored regression model (Liu & Wang, 2021).

The regression coefficient will provide an estimate of the size and direction of the treatment effect and its significance will be tested with a two-tailed 5% Type I error threshold. We will report our effect sizes as Hedge's g using unconditional (unadjusted) standard deviations accompanied by bootstrapped confidence intervals as a measure of uncertainty. Our secondary outcomes will be adjusted for multiple tests using Hochberg's step-up procedure (Menyhart et al., 2022).

Subgroup analyses

We will carry out analysis with sub-groups and include other variables of interest as predictors in regression analyses. In each case, the analytical model will be extended to include an interaction term, i.e. the interaction between subgrouping variable and trial arm. Given the higher probability of committing false statistical inferences due to multiple comparisons when analysing sub-groups, our primary and secondary outcomes will be adjusted for this using Hochberg's step-up procedure. We anticipate that most sub-group analyses will be

¹¹ We may adapt the models to include transformations such as a log link function depending on an assessment of skewness.

underpowered and will therefore be reported as exploratory analysis. Based on discussions with UpskillU and the Trial Steering Group, we will include analyses on:

- Age
- Sex
- Ethnicity (White British compared to other ethnic groups)
- Whether they have Special Educational Needs (SEND)
- Whether they currently have any mental health problems
- Whether they have an Education, Health or Care Plan (EHCP)
- Eligible for Free School Meals
- Care experienced
- Open to children's social care/early help
- Where they are in the youth justice system (i.e. cautioned, pre-court, post-court, custody etc.)
- Migrant/refugee status

As outlined in the secondary research questions, two key areas of this analysis will focus on CYP leaving custody, and CYP with special characteristics, particularly those from more marginalised groups (based on the variables listed above: girls and young women, CYP with mental health issues, migrant and refugee CYP, CYP from ethnic minority groups, care-experienced CYP, those with SEND and CYP who are economically disadvantaged). We will have access to this data from the referral forms filled in by CYP at the start of the programme, and ensuring the quality and completeness of this data will be key to this part of the analysis.

Further analyses

We will run robustness analyses for each outcome by including additional covariates that were imbalanced at baseline. If there are substantial changes in the trial delivery between the pilot and efficacy stages, we will conduct a sensitivity analysis of the results with and without pilot participants.

For the SDQ total difficulties score, we will report on the proportions of children and young people whose scores demonstrate reliable and clinically significant change (Wolpert et al., 2015) between baseline and follow-up, according to trial arm in order to assess individual level change and detect deterioration. We will compute:

- **Crossing of a Clinical Threshold (CCT)** on the measure, whereby the individual moves from the category of having a clinically significant problem to not having one, or vice versa. In our case we will set the clinical threshold to be ≥ 14 (above average or higher) based on the four-fold classification cited in the measure scoring guide. CCT will be reported on a group level as the proportion that crosses the threshold.

- **The Reliable Change Index (RCI)**, which estimates the amount of change required in an outcome measure to confidently conclude that the change observed is not solely attributable to measurement error. It is a function of the reliability of the scale, usually Cronbach alpha or test-retest reliability. Like CCT, RCIs are computed for individual cases, but will be reported for groups as the proportion having reliably changed.

We will also examine the correlation between the primary outcome (SRDS score) and police data on offending and violent offending to establish the level of difference (if at all) between self-reported and police reported measures at baseline and each follow-up stage. A similar analysis will be carried out for the JVQ and local police data on victimisation and exploitation. This will be an interesting comparison owing to the fact that CYP may under-report their own offending behaviour/ victimisation through self-reported measures, while on the other hand local police data may also under-represent behaviours which CYP are likely to self-report on.

We will also explore the variation of the gap between the date of randomisation and the first session descriptively by comparing it with demographic factors (including referral site) as well as baseline scores for the outcome measures. We will also look to explore the reasons behind this through interviews with UpskillU staff and young people as part of the implementation and process evaluation.

We recognise that CYP may lack trust in policing bodies and organisations that work with them, and will attempt to see if we can observe this through the data we collect. To this end, we will consider:

- stratifying our analysis of Social Support and Rejection Scale scores by referral partner to explore any differences in mentor-mentee relationship based on who the referral agency is (i.e. police, YOS, secure training centre or local authority team). We anticipate using an ordinal logistic regression for this analysis.
- comparing demographics from referral form data linked with local police data to establish any differences in convictions across ethnicity.

Longitudinal follow-up analyses

By the end of the final 12-month data collection period (July 2027) we would theoretically expect to have 18-month follow-up local police data on offending and violent offending available for at least 389 participants (accounting for 30% attrition). This is a sizeable sample and worth looking at as an exploratory, longer-term outcome given the length and nature of the intervention. Again, we will analyse this 18-month data using an ANCOVA model¹²,

¹² We may adapt the model to include transformations such as a log link function depending on an assessment of skewness.

including fixed effects in the regression model to account for referral partner, baseline offending levels and unbalanced predictor variables. We will also extend this to a mixed effects model using all 4 outcome time points (baseline, 6 months, 12 months and 18 months) and look at the time by treatment interaction. We anticipate that this analysis will be underpowered and will therefore be reported as exploratory.

Imbalance at baseline

We will report a baseline description of the trial participants using demographic data from UpskillU referral forms, referral partner monitoring data, and baseline outcome data. A list of the anticipated variables is included below. We will report descriptive statistics (means and standard deviations for continuous variables, percentages and counts for categorical variables) for each variable.

Table 6: Balance Testing Variables

Variable type	Variable
Demographic	Sex
Demographic	Age
Demographic	Ethnicity
Demographic	Disability
Demographic	English as an additional language
Demographic	Refugee/asylum-seeker status
Demographic	Eligibility for Free School Meals
Demographic	Special Education Needs (SEND)
Demographic	Looked-after status
Demographic	Open to/receiving CCE, CSE, CAMHS, Behavioural Support, Youth Justice Support
Demographic	Having an Education, Health or Care Plan (EHCP)

Process	Referral partner
Baseline outcome	Baseline SRDS score
Baseline outcome	Local police data on the number of criminal offences in the 12 months prior to baseline
Baseline outcome	Local police data on the number of violent offending in the 12 months prior to baseline
Baseline outcome	Baseline Module A JVQ score
Baseline outcome	Baseline local police data on number of incidents of victimisation and exploitation in the 12 months prior to baseline
Baseline outcome	Baseline total difficulties score on the SDQ
Baseline outcome	Baseline scores of the SSRS sub-scales

Missing data

We will assess all outcome data to explore whether the data is missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) and adjust our approach to analysis based on this assessment. Where data is MCAR, no imputation will be carried out and only available cases will be analysed. To assess MAR, we will carry out an additional analysis with the variables predictive of non-response, including variables not in the substantive model. We will use a multi-level logistic regression model with a binary outcome identifying when data is missing (=1) or not (=0). If only the outcome variables in the substantive model is MAR given covariates, those covariates will be included within the model for a more accurate interpretation. However, if a covariate in the substantive model is MAR given other covariates, analysis will be done after multiple imputation (MI) of that covariate using the method relevant to the substantive model (Predictive Mean Matching, Normal Imputation, etc.). Results from MI will be reported in addition to the headline impact estimates. About 10-20 iterations of the algorithm will be carried out to generate a minimum of 5 imputed datasets, and the analysis results on these datasets will be pooled using Rubin's Rules. These results will be subject to sensitivity analysis exploring the robustness of conclusions to departures from MAR, to produce estimates that adjust for missingness due to MNAR.

Compliance

A number of steps will be taken to monitor fidelity, compliance and identify risks to contamination. UpskillU coordinators will check all referral forms from referral partners to ensure the eligibility of CYP. Coram will receive referral forms for randomised CYP for quality assurance and to monitor the characteristics of those consenting and engaging in the trial. This will include characteristics such as age, sex, race and ethnicity, care status and sexuality. Qualitative interviews with referral partners and UpskillU staff will also be used to assess the referral process as part of the IPE.

Coram is responsible for randomisation and undertakes internal quality assurance checks to minimise any biases. Coram will also explore the influence of trial arm allocation compliance using Complier Average Causal Effect Analysis (CACE) (i.e. whether individuals in the intervention arm receive EXODUS), by including intervention receipt in an instrumental variable analysis. Compliance will be defined as attendance of at least 12 sessions of the EXODUS mentoring programme, i.e. that the CYP completes the intensive phase. For participants in the intervention arm, attendance at < 12 sessions will be considered non-compliant. We will compute CACE using instrumental variable regression by two-stage least squares to provide an estimate of the causal effect for EXODUS. The equation is as follows:

$$CACE = \frac{\bar{Y}_T - \bar{Y}_C}{(\bar{D} | Z = 1)}$$

where $\bar{Y}_T - \bar{Y}_C$ is the difference in mean outcomes for the two groups and $(\bar{D} | Z = 1)$ equals the observed treatment receipt rate for the treatment group. This framework comprises three variables: Y, the outcome measure; Z, which equals one for subjects randomized to treatment and zero otherwise; and D, which equals one for subjects who receive the treatment and zero otherwise.

As with the main analysis, these models will include fixed effects for referral partner, time from randomisation, and unbalanced baseline variables. We will do this for the primary outcome and compare these CACE coefficients with the intention-to-treat regression coefficients. We will report descriptive statistics on the rates of compliance, including rates of compliance by referral partner. Where possible, reasons for non-compliance will also be reported.

Fidelity

UpskillU has processes in place to ensure the fidelity of the intervention. Within 48 hours of each session, mentors complete and submit an online¹³ debrief form which includes:

- date, time, length and format of the session
- key discussions
- any documents completed
- any outcomes achieved
- any challenges, safeguarding, and health and safety information.

An UpskillU project coordinator is notified when these forms are completed. If a debrief form is not completed within 48 hours, a project coordinator will follow up with the mentor.

If a mentoring session is cancelled, a missed session debrief form is completed by the mentor detailing the reasons. If three consecutive sessions are missed, this is flagged to the project coordinator and the coordinator will take action to explore this.

An UpskillU project coordinator carries out random spot checks of mentor sessions and speaks to the CYP alone to check that the elements of the programme are working for them, including the relationship with the mentor and the venue. Additionally, a project coordinator has daily individual check-ins with mentors and weekly group meetings with mentors. Mentors also have access to coaching sessions with the Director of UpskillU as and when required.

For most of the original and extended phase, UpskillU collated debrief and missed session debrief forms into a spreadsheet to share with Coram as part of the monitoring data. Since July 2025, UpskillU has transitioned to using the RestorativU app to capture mentoring and monitoring data, and this data is now being received as a direct export from the app. The monitoring logs are being used to assess deviations according to key criteria against the **fidelity checklist**, as listed in **Table 3** under criteria 11, and below:

Dosage:

- at least 60% of organised in-person and virtual mentoring sessions are attended by CYP (yes/no)
- at least 50% of the planned COSA meetings are attended by CYP (yes/no) [Note: this is only relevant to full trial as the internal pilot does not run for a long enough time to capture COSA sessions which begin at around 38 weeks into the 12-month programme].

¹³ At the start of the programme, this was done on an online platform called “Typeform”, but over the last few months UpskillU have migrated this data capture process to the RestorativU app.

Coverage and consistency:

- whether four or more of the core themes¹⁴ are covered in the 12-session intensive phase by mentor (yes/no)
- risk assessments are completed for at least 80% of sessions by mentor (yes/no).

Coram assigns a fidelity score to each mentoring partnership that has had at least one session, and aggregated scores are used to report high, medium or low fidelity levels. This is being done at the end of the internal pilot and will be done at the end of the efficacy trial (if progression criteria are met). However, this data is being collected and quality assured at regular intervals, to ensure fidelity is being captured appropriately by UpskillU.

Coram will also separately report on the share of intervention YP that actually attend mentoring sessions out of the total that were randomised.

In the implementation and process evaluation, researchers who observe the intervention will also complete a fidelity checklist after each session observed.

Fidelity is also being explored in the implementation and process evaluation interviews, including whether CYP experience the different phases of the programme, whether the programme takes a restorative approach, whether the core themes are covered with CYP, whether prescribed activities take place and whether risk assessments were regularly undertaken and revisited.

These actions will help monitor fidelity and compliance as well as identify risks to contamination.

Presentation of outcomes

We will report our effect sizes as Hedge's g using unconditional (unadjusted) standard deviations accompanied by bootstrapped confidence intervals as a measure of uncertainty. The equation is given by:

$$ES = \frac{(T - C)_{adjusted}}{\delta_{pooled}}$$

Where T is the treatment mean, C is the control mean and δ_{pooled} is the pooled and weighted standard deviation across trial arms; this is calculated as:

¹⁴ 1. Self-identity, 2. Attitudes and values, 3. Risks and consequences, 4. Conflict and choices, 5. Relationships, influences and decisions and 6. Solutions and strategies.

$$\delta_{pooled} = \sqrt{\frac{(n_t - 1)\delta_t^2 + (n_c - 1)\delta_c^2}{n_t + n_c - 2}}$$

where n_t and n_c are the sample sizes for the trial arms, and δ_t^2 and δ_c^2 are the sample variances for each of the trial arms.

References

- Berry, V., Little, M., Axford, N., & Cusick, G. R. (2009). 'An evaluation of youth at Risk's coaching for communities programme', *The Howard Journal of Criminal Justice*, 48(1), 60-75.
- Creaney, S. (2018) 'Knife crime: how former offenders can make great mentors for at-risk teens', *The Conversation*. Available from: <https://theconversation.com/knife-crime-how-former-offenders-canmake-great-mentors-for-at-risk-teens-105880>, accessed 11 September 2023.
- Crown Prosecution Service (2022). 'Violent crime', Available at: <https://www.cps.gov.uk/crime-info/violent-crime>, accessed 16 July 2025.
- Flynn, S., Langdon, P., Hastings, R., Gray, K., Thompson, P., Coulman, E., Playle, R., Segrott, J., Lugg-Widger, F., & Moody, G. (2022). 'Evaluation protocol: Brief Solution Focused Therapy (BSFT) in 10–17-year-olds presenting at police custody: A Randomised Controlled Trial with internal pilot.' YEF, Available at: <https://youthendowmentfund.org.uk/wp-content/uploads/2022/09/YEF-Brief-Solution-Trial-Protocol-FINAL.pdf>, accessed 15 September 2023.
- Gaffney, H., Jolliffe, D. & White, H. (2022) 'Mentoring: Toolkit Technical Report', Youth Endowment Fund. Available from: https://youthendowmentfund.org.uk/wpcontent/uploads/2022/10/Mentoring-Technical-Report_Final.pdf. Accessed 24 April 2024.
- Goodman, R. (1997). 'The Strengths and Difficulties Questionnaire: A research note', *Journal of Child Psychology and Psychiatry*, 38: 581–586.
- Hamby, S. L., Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2004). The juvenile victimization questionnaire (JVQ): Administration and scoring manual. *Durham, NH: Crimes Against Children Research Center*, 10-18.
- Humayun, S., Herlitz, L., Chesnokov, M., Doolna, M., Landau, S., Scott S. (2017). 'Randomised controlled trial of Functional Family Therapy for offending and antisocial behaviour in UK Youth', *Journal of Child Psychology and Psychiatry*, 58(9), 1023-1032.
- Liu, Q., Wang, L. (2021). t-Test and ANOVA for data with ceiling and/or floor effects. *Behav Res* 53, 264–277. <https://doi.org/10.3758/s13428-020-01407-2>
- Menyhart O, Weltz B, Györfy B (2022). Correction: MultipleTesting.com: A tool for life science researchers for multiple hypothesis testing correction. *PLOS ONE* 17(9): e0274662. <https://doi.org/10.1371/journal.pone.0274662>

Smith, D. J., McVie, S., Woodward, R., Shute, J., Flint, J., & McAra, L. (2001). 'The Edinburgh study of youth transitions and crime: Key findings at ages 12 and 13', *British Journal of Criminology*, 43: 169-195.

Taylor, S., Blackshaw, E., Lawrence, H., Stern, D., Gilbert, L., Raghoo, N.. (2023) Randomised controlled trial of family group conferencing at pre-proceedings stage. Foundations. Available at: <https://foundations.org.uk/wp-content/uploads/2023/06/Randomised-controlled-trial-family-group-conferencing.pdf>

Tolan, P. H., Henry, D. B., Schoeny, M. S., Lovegrove, P., & Nichols, E. (2014). 'Mentoring programs to affect delinquency and associated outcomes of youth at risk: A comprehensive meta-analytic review', *Journal of experimental criminology*, 10: 179-206.

Walters, S.J., Jacques, R.M., dos Anjos Henriques-Cadby, I.B. *et al.* (2019) 'Sample size estimation for randomised controlled trials with repeated assessment of patient-reported outcomes: what correlation between baseline and follow-up outcomes should we assume?', *Trials*, 20(566). <https://doi.org/10.1186/s13063-019-3671-2>

Williams, A. (2019) 'Family support services delivered using a restorative approach: A framework for relationship and strengths-based whole-family practice,' *Child & Family Social Work*, 24(4): 555-564.

Wolpert, M., Görzig, A., Deighton, J., Fugard, A.J.B., Newman, R. and Ford, T. (2015), Comparison of indices of clinically meaningful change in child and adolescent mental health services: difference scores, reliable change, crossing clinical thresholds and 'added value' – an exploration using parent rated scores on the SDQ. *Child Adolesc Ment Health*, 20: 94-101. <https://doi.org/10.1111/camh.12080>

Zand, D. H., Thomson, N., Cervantes, R., Espiritu, R., Klagholz, D., LaBlanc, L. and Taylor, A. (2009). 'The mentor–youth alliance: The role of mentoring relationships in promoting youth competence', *Journal of Adolescence*, 32(1): 1-17



youthendowmentfund.org.uk



hello@youthendowmentfund.org.uk



[@YouthEndowFund](https://twitter.com/YouthEndowFund)

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413
