STATISTICAL ANALYSIS PLAN

The Summer Jobs Programme – A randomised multi-site trial to evaluate an employment intervention aiming to enhance positive outcomes for young people at risk of violence

IFF Research

Principal investigator(s): Kelsey Beninger, Sashka Dimova

youth futures FOUNDATION

YOUTH ENDOWMENT FUND

Summer Jobs Programme: Statistical analysis plan

Project title	Summer Jobs Programme – a randomised multi-site control trial to evaluate an employment intervention aiming to enhance positive outcomes for young people at risk of violence.
Developer (Institution)	UK Youth
Evaluator (Institution)	IFF Research
Principal investigator(s)	Kelsey Beninger, Sashka Dimova
SAP author(s)	Sashka Dimova
Trial design	Two-armed, muti-site randomised control trial, with randomisation at the individual young person level
Trial type	Efficacy Study
Evaluation setting	Community (England and Wales)
Target group	16- to 20-year-olds at risk of violence
Number of participants	**Planned recruitment (YEF target):** 3,000 young people in total ○ 1,200 in the Year 2 Internal Pilot (2025) across 18 Local Delivery Partners (LDP) ○ 1,800 in the Year 3 Efficacy Study (2026) across 30 LDPs during Year 3 **Actual recruitment following the Year 2 Internal Pilot:**

	• 787 participants randomised in Year 2 (vs the target of 1,200) across 18 LDPs • Revised target of 1,650 participants in Year 3 (revised projection based on Year 2 delivery) across 23 LDPs[1]
Primary outcome and data source	The two co-primary outcomes, measured using national administrative data, are: Offending (all young people in England and Wales): • Source: DfE–PNC Datashare Employment, Education and Training (EET) (young people in England only): • Source: LEO dataset (employment or education/training records)
Secondary outcome and data source	The four secondary outcomes, measured for young people in England and Wales through a self-report online questionnaire are: • Strength and Difficulties Questionnaire (SDQ)[2] total difficulties score, 12 to 16 weeks after randomisation; • Whether young person reports being in Employment, Education or Training, 12 to 16 weeks after randomisation • New Philanthropy Capital (NPC) Journey to Employment (JET) Framework questionnaires[3]

[1] The Year 2 pilot recruited 787 participants across 18 LDPs, below the original target of 1,200. Following the pilot, it was agreed to revise the target for the efficacy study to 1,650 participants across 23 LDPs, reflecting a more realistic recruitment goal rather than a formal power calculation. Importantly, the study remains well-powered to detect small, policy-relevant effects. Power calculations using the revised sample indicate a modest increase in the minimum detectable effect size (MDES), with detectable effects in the range of approximately 0.14 to 0.19 standard deviations (Cohen's d) across the co-primary outcomes. The decision to proceed with the Year 3 evaluation reflects the assessment that the revised sample size would not materially compromise statistical power or the robustness of the study. According to Youth Endowment Fund Magnifying Glass guidance, this level of precision corresponds to an MG 4 rating (MDES 0.11–0.19).

[2] For more information see Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345.

[3] Copps, J. and Plimmer, D. (2013). The Journey to Employment (JET) Framework: Outcomes and tools to measure what happens on young people's journey to employment. [Available online:

	aspiration to work score, 12 to 16 weeks after randomisation • New General Self-Efficacy Scale (NGSC),[4] total scale score, 12 to 16 weeks after randomisation

SAP version history

Version	Date	Changes made and reason for revision
5.0	29.05.2026	Updated to include stratification by borough in randomisation and analysis procedures; clarified that no manual manipulation of participants occurs after randomisation
4.0	20.04.2026	Final updates made in response to third round of peer review
3.0	26.03.2026	Further updates made in response to second round of peer review after initial updates
2.0	12.03.2026	Updates made in response to peer review
1.0 [*original*]	26.01.2026	Efficacy Trial set-up

Any changes to the design or methods need to be discussed with the YEF Evaluation Manager and the developer team prior to any change(s) being finalised. Describe in the table above any agreed changes made to the evaluation design. Please ensure that these changes are also reflected in the SAP (CONSORT 3b, 6b).

https://npproduction.wpenginepowered.com/wpcontent/uploads/2018/07/JET-framework-FINAL-Jan-2015.pdf

[4] Chen, G., Gully, S. M., & Eden, D. (2001). *New General Self-Efficacy Scale (NGSE)* [Database record]. APA PsycTests. https://doi.org/10.1037/t08800-000

Table of contents

SAP version history 3

Table of contents 4

Introduction 5

Design overview 6

Sample size calculations overview 8

Analysis 15

Appendix A 4

Introduction

This Statistical Analysis Plan (SAP) sets out the predefined analytical methods for evaluating the Summer Jobs Programme, a six-week youth employment intervention delivered by UK Youth through Local Delivery Partners (LDPs). The programme provides one week of paid work-readiness training followed by five weeks of paid employment for young people aged 16 – 20 who are at risk of violence[5]. Its overarching aim is to reduce offending and improve engagement in education, employment, and training (EET).

The evaluation is a two-arm, multi-site randomised controlled trial with individual-level randomisation stratified by LDP. Eligible and consenting young people who complete right-to-work checks are allocated 1:1 to either the intervention group (offered Summer Jobs) or the control group (receiving business-as-usual support). The trial consists of a Year 2 Internal Pilot Study (2025) followed by a Year 3 Efficacy Study (2026). The original combined target for Years 2 and 3 was 3,000 participants, with 1,200 planned for Year 2 and 1,800 for Year 3. The Year 2 pilot recruited 787 participants across 18 LDPs, below the planned 1,200.

Following the pilot, the Year 3 target was revised to 1,650 participants across 23 LDPs. This revised target reflects a more realistic and achievable recruitment goal based on Year 2 delivery, while ensuring that the combined planned sample of 2,434 participants remains well-powered to detect, small policy-relevant effects on the co-primary outcomes. Updated power calculations indicate that the modest increase in minimum detectable effect size (MDES) corresponds to a Cohen's d of approximately 0.14–0.19. According to Youth Endowment Fund Magnifying Glass guidance, this level of precision corresponds to a high-confidence rating (MG 4), whereas an MG 5 rating would require an MDES of 0.10 or below. The revision therefore balances practical delivery considerations with the methodological rigor required to generate robust, policy-relevant evidence.

The purpose of the analyses described in this SAP is to estimate the impact of being offered Summer Jobs on prespecified co-primary outcomes (offending and EET status), secondary survey-based outcomes, and exploratory subgroup effects, using an intention-to-treat framework. This document outlines the statistical models, handling of missing data, adjustments for multiplicity, and planned sensitivity and robustness checks.

A full description of the intervention, implementation model, trial procedures, and outcome measures is available in the Summer Jobs Trial Protocol.[6] This SAP should be read in

[5] 'At risk of violence' is defined in the protocol; eligible participants meet general criteria (age, UK work eligibility, programme availability) and at least one risk indicator such as prior contact with Youth Offending Teams, school exclusions, being in care, or having experienced or been exposed to violence (see protocol, Participants section).

[6] See the following sections of the protocol: 'intervention' (pp. 13-20), 'implementation and process evaluation' (pp. 51-55), 'outcome measures' (pp. 39-50).

conjunction with that protocol, which provides complementary operational and methodological detail.

Design overview

Trial design, including number of arms		Two-arm, multi-site randomised controlled trial
Unit of randomisation		Individual young person
Stratification variables (if applicable)		Local Delivery Partners (LDPs) and Borough
Co-primary outcomes	variable	1. Offending 2. Employment, Education and Training (EET)
	measure (instrument, scale, source)	1. Offences linked to criminal records, convictions and cautions at 6 months after randomisation (1 if young person has a recorded criminal offence,[7] 0 otherwise; the Police National Computer dataset) 2. Being in Employment, Education or Training at 6 months after randomisation (1 if young person classes as 'EET', 0 otherwise; classes as 'EET' if employed or in education and training using the Longitudinal Education Outcome dataset)
Secondary outcome(s)	variable(s)	1. Behavioural, emotional and relationship problems 2. Self-reported EET 3. Aspiration to employment 4. Self-efficacy
	measure(s)	1. Strength and Difficulties Questionnaire (SDQ)[8] total difficulties score, young-person self-report fielded in a survey 12 to 16 weeks after randomisation

[7] Being given a caution does not mean you are "non guilty". Therefore, we can use both cautions and convictions in our analysis, but we should not interpret the former as "non guilty" offences.
[8] For more information see Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345.

	(instrument, scale, source)	2. Young person self-report on whether is in Employment, Education or Training fielded in a survey 12 to 16 weeks after randomisation 3. New Philanthropy Capital (NPC) Journey to Employment (JET) Framework questionnaires[9] aspiration to work score, young-person self-report fielded in a survey 12 to 16 weeks after randomisation 4. The New General Self-Efficacy Scale (NGSC),[10] total scale score, young-person self-report fielded in a survey 12 to 16 weeks after randomisation
Baseline for primary outcome	**variable**	1. Offending 2. Employment, Education and Training (EET)
	measure (instrument, scale, source)	1. Offences linked to criminal records (1 if young person has a criminal record,[11] 0 otherwise; DfE-PNC Datashare) prior to randomisation 2. Being in Employment, Education or Training (1 if young person classes as 'EET', 0 otherwise; classes as 'EET' if employed or in education and training using the Longitudinal Education Outcome dataset) in the four weeks prior to randomisation
Baseline for secondary outcome	**variable**	1. Behavioural, emotional and relationship problems 2. Self-reported EET 3. Aspiration to employment 4. Self-efficacy
	measure (instrument, scale, source)	1. Strength and Difficulties Questionnaire (SDQ)[12] total difficulties score, young-person self-report fielded in a survey prior to randomisation 2. Self-reported whether young person is in Employment, Education and Training, young person self-report fielded in a survey prior to randomisation

[9] Copps, J. and Plimmer, D. (2013). The Journey to Employment (JET) Framework: Outcomes and tools to measure what happens on young people's journey to employment. [Available online: https://npproduction.wpenginepowered.com/wpcontent/uploads/2018/07/JET-framework-FINAL-Jan-2015.pdf

[10] Chen, G., Gully, S. M., & Eden, D. (2001). *New General Self-Efficacy Scale (NGSE)* [Database record]. APA PsycTests. https://doi.org/10.1037/t08800-000

[11] Being given a caution does not mean you are "non guilty". Therefore, we can use both cautions and convictions in our analysis, but we should not interpret the former as "non guilty" offences.

[12] For more information see Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345.

		3. NPC Journey to Employment (JET) Framework questionnaires[13] aspiration to work score, young-person self-report fielded in a survey prior to randomisation 4. The New General Self-Efficacy Scale (NGSC),[14] total scale score, young-person self-report fielded in a prior to randomisation

Sample size calculations overview

We present minimum detectable effects (MDEs) in percentage points, risk ratios, and the associated Cohen's d effect size for offending and EET status. The study is powered for the co-primary outcomes only; secondary outcomes and subgroup analyses described in the protocol are exploratory and the study is not powered to detect differences within subgroups.

Sample size estimations for offending are based on the total number of participants from the Year 2 Internal Pilot Study (2025) and Year 3 Efficacy Study (2026). A total of 2,432 young people will be included across both years (782 in Year 2 and 1,650 in Year 3).

The analysis on the EET outcome will focus on young people in England because data will be obtained via LEO, which only covers young people educated in England. Sample size estimations are based on the actual Year 2 Internal Pilot Study (2025) recruitment and the revised Year 3 Efficacy Study (2026) targets.

- Offending outcome: 787 participants were randomised, and 5 withdrew formally from the evaluation, resulting in actual 782 Year 2 participants; plus 1,650 participants in Year 3 (total 2,432).
- EET outcome (England only): 747 participants in Year 2 plus estimated 1,580 participants in Year 3 (total 2,327).

Power calculations account for attrition and assume an equal number of young people in the intervention and control groups. Attrition is defined as participants who cannot be matched in administrative records or who formally withdraw from the study. Attrition rates of 10% (baseline and follow-up data available for 2,188 young people for offending and 2,094 for EET) and 20% (baseline and follow-up data available for 1,946 young people for offending and

[13] Copps, J. and Plimmer, D. (2013). The Journey to Employment (JET) Framework: Outcomes and tools to measure what happens on young people's journey to employment. [Available online: https://npproduction.wpenginepowered.com/wpcontent/uploads/2018/07/JET-framework-FINAL-Jan-2015.pdf
[14] Chen, G., Gully, S. M., & Eden, D. (2001). *New General Self-Efficacy Scale (NGSE)* [Database record]. APA PsycTests. https://doi.org/10.1037/t08800-000

1,862 for EET) are considered. We assume it is reasonable that attrition will not exceed 20%, as we rely on administrative data for the co-primary analysis.

Given that there are two primary outcomes, we adjust the significance level to account for multiple comparisons using a Westfall-Young correction rather than the more conservative Bonferroni correction. Since this correction involves resampling methods it is approximated in the MDE calculations in R by slightly adjusting the alpha level used in power functions.[15]

Power calculations assume two-tailed testing and a power level of 0.80. Calculations were run in R (code in Appendix 3) and will be updated following randomisation in June 2026, once final sample sizes are confirmed.

Offending (co-primary outcome)

There is limited existing evidence on the proportion of young people in our target population who will have committed an offence 6 months after randomisation. The study population is more diverse and includes young people at risk of violence, whereas most existing research focuses on individuals who are already known to the criminal justice system. Figures from previous studies indicate that different risk factors lead to higher or lower risk of offending.[16] To reflect this, we present power calculations for two different assumed incidence rates of offending. We assume that 30% to 40%[17] of young people in the control group[18] will be offenders. This aligns with evidence indicating that pupils who experienced some form of alternative provision (AP) in England in the last decade and who were linked to the Police National Computer (PNC), around 40% had some criminal record[19]. However, it is uncertain how representative this assumption is for our target population. To address this, we intend to use PNC data for the population involved in the Year 2 Pilot study to cross-check and validate our assumption regarding the baseline offending rate. We submitted an application to receive PNC data for the pilot population, which is currently under review. The application is expected to be processed by September 2026.

[15] We approximate the adjustment by reducing the nominal alpha from 0.05 to 0.03 in the R power calculations

[16] There are roughly around 58 individual-level, family and school level predictors of crime. Some risk factors associated with higher risk of offending include mental health problems, substance use, adverse childhood experience, parent maltreatment and neglect, attendance and attainment. For more information on the risks and protective factors of youth crime see Risk and protective factors of youth crime: An umbrella review of systematic reviews and meta-analyses - ScienceDirect

[17] This figure is grounded in the analysis of the DfE-MoJ dataset that suggests that, among pupils who have been in Alternative Provision in the last 10 years in England, 30% have some violent criminal record and 50% have some criminal record.

[18] This reflects plausible baseline levels of offending for young people at risk of violence and forms the basis for estimating the study's minimum detectable effect. This does not imply that uncertainty differs between groups; the intervention group's rate is unknown and expected to vary according to programme effects.

[19] Education, children's social care and offending

Co-primary outcome: Offending		Protocol (Incidence rate in control group =30%)	Protocol (Incidence rate in control group =40%)
Minimum Detectable Effect Size (MDES) in percentage points		5.44	5.90
Risk Ratio		0.819	0.853
Cohen's D Effect Size approximation[20]		-0.152	-0.140
Pre-test/ post-test correlations	level 1 (participant)	N/A	
	level 2 (cluster)	N/A	
Intracluster correlations (ICCs)	level 1 (participant)	N/A	
	level 3 (cluster)	N/A	
Alpha[21]		0.03	
Power		0.8	
One-sided or two-sided?		Two-sided	
Average cluster size		N/A	
Number of clusters	intervention	N/A	

[20] To calculate Cohen's D from the odds ratio, we take the natural logarithm of the odds ratio and divide by 1.81. See Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(22), 3127–3131. https://doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

[21] We set the significance level (alpha) at 0.03 rather than the conventional 0.05 to account for having two co-primary outcomes. This approximates a correction for multiple comparisons, using a Westfall-Young approach rather than the more conservative Bonferroni method.

Co-primary outcome: Offending		Protocol (Incidence rate in control group =30%)	Protocol (Incidence rate in control group =40%)
	control	N/A	
	total	N/A	
Number of participants	intervention	1,216	
	control	1,216	
	total	2,432	

For offending, based on a total sample of 2,432 young people across Years 2 and 3, the study will be able to detect a minimum difference of approximately 5.4 to 5.9 percentage points[22], depending on the assumed control group offending rate (30–40%).The relative reduction of 5.4 to 5.9 percentage points corresponds to around 15 to 22% relative reduction. These differences correspond to risk ratios ranging from approximately 0.82 to 0.85 and standardised effect sizes (Cohen's d) in the range of −0.15 to −0.14, indicating reductions in offending in the intervention group relative to the control group.

This suggests that the multi-site trial will be well-powered, as evidence from the USA shows similar Summer Jobs programmes have reduced violence by 30% to 43%.[23,24] However, those studies focus specifically on violent crime, whereas our measure includes all types of offending. In addition, levels and patterns of violence may differ between the US and England, and the programme tested in this study has been adapted to the English context.

[22] Note on interpretation: The MDES is expressed in absolute percentage points, which represents the direct difference between control and intervention groups detectible in the study. To compare with US evidence that reports percent reductions (e.g., Heller, 2014; Modestino, 2019), the relative reduction can be calculated by dividing the absolute difference by the assumed control group rate. For example, a 5.4 pp reduction with a 30% control rate corresponds to an 18% relative reduction. So the relative reduction range that is detectible is approximately 14% – 22%.

[23] Heller, S. B. (2014). Summer jobs reduce violence among disadvantaged youth. Science, 346(6214), 1219–1223. https://doi.org/10.1126/science.1257809.

[24] Modestino, A. S. (2019b). How do summer youth employment programs improve criminal justice outcomes, and for whom? Journal of Policy Analysis and Management, 38(3), 600–628.

EET (co-primary outcome, England only)

It is well documented that young people who are at risk of being offenders are more likely to end up being NEET.[25] However, it is challenging to determine the exact figure for the proportion of 16–20-year-olds at high risk of offending who end up being NEET due to limited specific data. There is evidence suggesting that around 60% of young people who had spent time in AP were NEET at age 19, compared to just 12% of their mainstream peers.[26] Analysis of Youth Offending Service data found that approximately 30% of young offenders 16 to 24 were NEET prior to offending.[27] Considering the uncertainties, we present power calculations for three different incidence rates. We assume that 25%, 35% and 50 % of young people in the control group are NEET, which equates to 75%, 65% and 50% EET rates at 6 months post-randomisation.

Co-primary outcome: EET Status		**Protocol (Incidence rate in control group =50%)**	**Protocol (Incidence rate in control group =65%)**	**Protocol (Incidence rate in control group =75%)**
Minimum Detectable Effect Size (MDES) in percentage points		6.22%	5.82%	5.20%
Risk Ratio		1.124	1.090	1.069
Cohen's D Effect Size approximation[28]		0.138	0.148	0.166
Pre-test/ post-test correlations	level 1 (participant)	N/A		
	level 2 (cluster)	N/A		

[25] Youth Justice Board, National Indicator 45: Education Training and Employment data YJB Corporate Brochure - Education (English)
[26] Integrated (2020). Fewer Exclusions. Better Alternative Provision. See Integrated Annual Report 2020

[27] Knight, V. (2020). *Scaling it Down? A Study to Identify How 16+ Young Offenders Across Leicestershire Engage with Education, Training or Employment*. **British Journal of Community Justice.**

[28] To calculate Cohen's D from the odds ratio, we take the natural logarithm of the odds ratio and divide by 1.81. See Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(22), 3127–3131. https://doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

Co-primary outcome: EET Status		Protocol (Incidence rate in control group =50%)	Protocol (Incidence rate in control group =65%)	Protocol (Incidence rate in control group =75%)
Intracluster correlations (ICCs)	level 1 (participant)	N/A		
	level 3 (cluster)	N/A		
Alpha[29]		0.03		
Power		0.8		
One-sided or two-sided?		Two-sided		
Average cluster size		N/A		
Number of clusters	intervention	N/A		
	control	N/A		
	total	N/A		
Number of participants	intervention	1,164		
	control	1,163		
	total	2,327		

For EET status, based on a total sample of 2,327 young people educated in England across Years 2 and 3, the study will be able to detect a minimum difference of approximately 5.2 to

[29] We set the significance level (alpha) at 0.03 rather than the conventional 0.05 to account for having two co-primary outcomes. This approximates a correction for multiple comparisons, using a Westfall-Young approach rather than the more conservative Bonferroni method.

6.2 percentage points, depending on the assumed control group EET rate (50–75%).These differences correspond to risk ratios ranging from approximately 1.07 to 1.12, with associated standardised effect sizes (Cohen's d) in the range of 0.14 to 0.17, indicating higher rates of EET among young people receiving the intervention.

This difference is realistic based on similar contexts, such as DWP's Kickstart programme, where a net impact of 11 percentage points in competitive employment was observed.[30] However, evidence on the impact of Summer Jobs on employment and education outcomes is limited and mixed, with U.S. studies showing no increase in earnings, employment, or education.[31]

The study is considered well-powered for the EET outcome because the sample size allows detection of minimum differences in the range of 5.2–6.2 percentage points, corresponding to standardised effect sizes of 0.14–0.17. According to YEF guidance, these effect sizes are sufficient to achieve a high-confidence MG 4 rating[32].

Key differences between the original and updated estimates

Based on the actual Year 2 internal pilot sample and the revised Year 3 recruitment targets, we have updated the minimum detectable effect sizes (MDE), and Cohen's d for the co-primary outcomes. For offending, the study can detect differences of approximately 5.4 to 5.9 percentage points, corresponding to Cohen's d of −0.15 to −0.14, indicating reductions in offending in the intervention group. For EET status, detectable differences range from approximately 5.2 to 6.2 percentage points, and Cohen's d of 0.14–0.17, reflecting higher rates of EET among intervention participants.

While there is a modest increase in the detectable effect sizes compared with the original calculations, the study remains adequately powered to detect small but policy-relevant effects, consistent with effect sizes typically observed in comparable social intervention trials.[33,34,35]

---

[30] See page 40 at Davis, J. M. V., & Heller, S.B. (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *The Review of Economics and Statistics*, *102*(4), 664–677. https://doi.org/10.1162/rest_a_00850.

[31] Gelber, A., Isen, A. & Kessler, J.B. (2016). The effects of youth employment: Evidence from New York City lotteries. *The Quarterly Journal of Economics*, 131(1), 423-460. https://doi.org/10.1093/qje/qjv034.

[32] See YEF-Magnifying-Glass-Guidance.pdf

[33] Heller, S. B. (2014). Summer jobs reduce violence among disadvantaged youth. Science, 346(6214), 1219–1223. https://doi.org/10.1126/science.1257809.

[34] Modestino, A. S. (2019b). How do summer youth employment programs improve criminal justice outcomes, and for whom? Journal of Policy Analysis and Management, 38(3), 600–628.

[35] Davis, J. M. V., & Heller, S.B. (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *The Review of Economics and Statistics*, *102*(4), 664–677. https://doi.org/10.1162/rest_a_00850.

Outcome	Control Rate	Original MDE (%)	Original Cohen's d	Updated MDE (%)	Updated Cohen's d
Offending	30–40%	5.2–5.4	−0.08	5.4–5.9	−0.15- −0.14
EET	50–75%	4.8–5.8	0.08	5.2-6.2	0.14–0.19

Analysis

The outcome analysis will follow an intention-to-treat (ITT) approach, meaning all participants will be included in the groups they were randomly assigned to, regardless of whether they received the treatment, withdrew from Summer Jobs, or experienced deviations in programme implementation. This conservative approach captures the average effect of offering Summer Jobs compared to business as usual, regardless of participants' adherence to the assigned group. Analyses will be conducted using R and all analytic choices, including model specification and covariate adjustment are pre-specified prior to data collection, with syntax provided in technical appendix.

The following descriptive statistics will be reported:

1. Baseline demographic characteristics on key variables (sex at birth, age, ethnicity, special education needs, referral source)[36] will be described, using descriptive statistics (means, standard deviation, numbers, and percentages) for the overall sample, and for the intervention and control groups.
2. The balance between the intervention and control group at baseline, comparing the size of differences in effect sizes on demographic characteristics (sex at birth, age, ethnicity, special education needs) and on the primary and secondary outcomes. Standardised mean differences (SMDs) will be calculated for each variable to provide a scale-free measure of imbalance between groups (discussed further in **the Imbalance at baseline** section).

[36] Certain additional variables such as Children in Need (CiN), Children Looked After (CLA), and free school meal (FSM) eligibility will not be included. CiN and CLA are special-category personal data requiring additional approvals, which can be challenging and time-consuming to obtain. FSM data is less relevant for the 16–19-year-old target population and is not routinely collected for many participants.

3. The distribution of primary and secondary outcomes at follow-up will be presented as plots for both the intervention and control groups. Descriptive summary statistics (mean, SD, median, IQR for continuous outcomes; counts/proportions for categorical outcomes) will accompany these plots.
4. Attrition rates will be reported for both the intervention and control across all primary and secondary outcomes.
5. Sample size at baseline and follow-up survey completion will be noted for both the intervention and control group.

Primary outcome analysis

To estimate the impact on the co-primary outcomes (i.e. binary indicator on offending up to 6 months post-randomisation; and EET status at 9 months post randomisation) we will use a logistic regression model.

Data from the Year 2 Internal Pilot Study (2025) and Year 3 Efficacy Study (2026) and from all Local Delivery Partners will be pooled, and the model will include a fixed intercept for each LDP and borough to account for stratification. By including LDP and boroughas fixed effects, we ensure that any systematic differences between LDPs are adequately controlled for in the model. The full model is as follows:

$$logit(P(Y_{ij} = 1)) = \beta_0 + \beta_1 SJ_{ij} + \beta_2 B_i + \sum_{k=1}^{K-1} \gamma_k LDP_{jk} + \sum_{k=1}^{K-1} \gamma_k Borough_{jk} + \beta_3 Phase_i$$

- Y_{ij} is the outcome (offending or EET) for young person i referred by LDP j
- SJ_{ij} is a binary indicator denoting whether a young person is assigned to the intervention or the control group
- B_i represents the baseline rate for offending and EET
- $Phase_i$ represents a binary indicator for trial phase (1 = Year 3 Efficacy Study, 0 = Year 2 Internal Pilot Study).
- LDP_{jk} and $Borough_{jk}$ is a series of binary indicators represnting each LDP (fixed effects) reflecting that this is a multi-site trial

Given that there are two primary outcomes, appropriate adjustments will be made to the confidence intervals and p-values following the Westfall-Young correction. Both the p-values and statistical significance level will be reported. P-values will be reported as continuous probabilities, with no dichotomisation or use of asterisks. Standard errors for the logistic regression coefficients will be estimated at the model level, as provided by the maximum likelihood estimation of the logistic regression. LDP and brough fixed effects are included to account for systematic differences between sites/boroughs. Odds and relative risk ratio for the primary outcome analysis will be reported.

Data for offending at 6-month post randomisation is anticipated to be available in March 2028, and analysis will be conducted by March 2029.

The analysis of EET outcomes is subject to the availability of data from the LEO dataset. It is important to note that the publication and availability of LEO data are subject to variability, both in terms of when the data are collected and when they become accessible for research purposes. As such, the timelines for accessing EET outcome data should be regarded as best estimates and may be subject to change.

Based on current guidance from DfE, we have been advised to allow for a lag of approximately two years following the academic year in which the data are generated before it becomes available for analysis.

Given this lag, we anticipate that data relating to EET outcomes at 6 and 9 months post-randomisation will become available by March 2029. Subject to data access timelines holding as expected, we plan to complete the analysis of these outcomes by June 2029.

Secondary outcome analysis

For secondary outcomes that are continuous variables, a linear regression model with fixed effects will be applied, similar in structure with the logistic model, but with a continuous outcome variable. Fixed effects will include Local Delivery Partner (LDP) indicators and trial phase, ensuring that any systematic differences across sites or phases are controlled for. For ordinal or multinomial secondary outcomes, a generalised linear model will be used, assuming the appropriate distribution for the data.

Secondary outcomes are considered exploratory, and the study is not powered for these analyses. For transparency, both unadjusted and multiplicity-adjusted p-values will be reported, although inference will focus primarily on the co-primary outcomes.

Subgroup analyses

The study will not be powered for sub-groups analysis. However, exploratory analyses will be undertaken on the co-primary outcomes for several key subgroups, despite the study not being statistically powered for these analyses. These analyses were pre-specified in the evaluation protocol, and the subgroup definitions presented here are aligned with those described in the protocol.These subgroup analyses are intended to provide additional insights into potential heterogeneity of treatment effects across different demographic groups, helping to explore how the intervention may affect different populations differently.

We will undertake the following subgroup analysis:

- Sex at birth (two categories male vs female)
- Age (two categories based on age bands 16-18; and 18-20)

- Ethnicity (five categories i.e. White; Mixed or Multiple ethic group; Asian or Asian British; Black, Black British, Caribbean or African, Other) to understand any differences in the outcomes for the key demographic groups.
- Whether young person has special educational needs
- Type of referral (i.e. external agency/ internal LDP network) to assess difference in outcome by referral source
- English versus Welsh sample (relevant only for offending).

A sub-group analysis will be conducted to estimate the conditional average treatment effects (CATE) by using interaction terms for all co-primary outcomes. The analysis will be conducted using a single equation model, where the interaction indicator is interacted with the subgroup variable. A separate regression model will be estimated for each subgroup dimension (e.g., sex, age, ethnicity). This allows us to estimate how the interaction effect differs for each subgroup, capturing any potential heterogeneity in the treatment response. This means that dummy variables for the sub-groups will be interacted with the binary treatment variable.

Further analyses

We will run additional secondary specifications to test the robustness of the results. In expectation, due to the randomisation of treatment, doing so should not change the point estimates of impact, but may change the confidence intervals.

Specifically, we will estimate a regression without covariates to test robustness of the results, controlling only for the treatment assignment as covariate.

Longitudinal follow-up analyses

Long-term follow up analysis will be conducted for offending using the same outcome measures obtained via the PNC and using the same specification as the primary analysis at 6 months post randomisation. Specifically, the same model type (logistic regression for the binary outcome) will be used to ensure consistency with the primary analysis. Offending will be measured at 12 months after the end of the programme. It is expected that data for offending will be available between October and December 2028, and it is anticipated that that the long-term analysis will be undertaken by March 2029. To ensure coherence in reporting and to align with the availability of data from LEO, this analysis will be published alongside the primary analysis of EET outcomes at 6 and 9 months post-randomisation. The EET data are expected to be available by March 2029, with analysis completed by June 2029.

Both analyses—the long-term offending and the primary and long-term EET outcomes—will therefore be included in an addendum report drafted by June 2029 and published by November 2029.

In summary:

Summer Jobs administrative data access and analysis plan

		6m post-randomisation (offending) 6m post-randomisation (EET)			12m post-intervention end (offending) 9m post-randomisation (EET)		
Evaluation	**Intervention delivery**	**Application:**	**Access & analysis:**	**Report analysis included in**	**Application: 12m data**	**Access & analysis (latest): 12m data**	**Report analysis included in**
Internal Pilot Year 2 (2025).	Late July to Sept 2025	**PNC:** April 2025 **LEO:** N/A	**PNC:** September 2026 **LEO:** N/A	March 2028	**PNC:** Sept 2026 **LEO**: N/A	**PNC:** Sept 2027 **LEO:** N/A	PNC: March 2029 **LEO:** N/A
Efficacy Study Year 3 (2026).	Late July to Sept 2026	**PNC:** April 2026 **LEO:** June 2026	**PNC:** March 2028 **LEO:** March-June 2029[37]	**PNC:** March 2028 **LEO:** June 2029	**PNC:** Sept 2027 **LEO**: June 2026	**PNC:** Jan-March 2029 **LEO:** March-June 2029	**PNC:** June 2029 **LEO:** June 2029

[37] The publication and availability of LEO data are subject to variability, both in terms of when the data are collected and when they become accessible for research purposes. As such, the timelines for accessing EET outcome data should be regarded as best estimates and may be subject to change.

Imbalance at baseline

We will present a table of baseline balance between treated and control groups, for all individuals recruited in the groups that were randomised and for the sample of individuals with endline outcome data. This will report means and standard deviations for continuous variables (and only means for binary outcomes). Data sources will include: management information data providing baseline characteristics, surveys and administrative data on outcome measures. The basic characteristics to be presented include baseline demographic characteristics of the young person (sex at birth, age, ethnicity, special educational needs, and referral source), and baseline measures of all co-primary and secondary outcomes, including offending, EET status, behavioural and emotional difficulties (measured using the Strengths and Difficulties Questionnaire), self-reported EET status, aspirations to employment, and self-efficacy.

Missing data

If there is less than 5% missingness overall (i.e., the primary analysis model includes at least 95% of randomised young people), a complete-case analysis will be carried out. The number of complete cases included in the analysis will be explicitly reported.

If there is more than 5% missingness overall, analysis will be undertaken to understand if the data appears to be missing completely at random (MCAR), or whether the weaker Missing at Random (MAR) assumption applies. The extent of missingness will be assessed, and then also explore the pattern of any identified missingness. To assess whether there are systematic differences between those who provide data and those who do not and thus whether these factors should be included in analysis – missingness will be modelled through a logistic regression model at follow-up as a function of baseline covariates, including indicator denoting whether a young person is allocated to the intervention or control group. The analysis model for this approach will mirror the model given above but the outcome will be a binary variable identifying missingness (yes/no).

If MAR is deemed plausible, a full-information maximum likelihood (FIML) approach will be applied to account for missing values. FIML can be estimated in a single model and simulation studies show that it can reduce bias as well as MI.[38] If MAR does not hold, the potential limitations and risk of bias due to missingness will be explicitly described in reporting. Sensitivity analysis will be conducted using best-case and worst-case scenarios for missing

[38] Multiple imputation is not necessarily the gold standard in missing data handling in RCTs, with other (simpler) methods providing similarly unbiased estimates: Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials?. Statistical methods in medical research, 27(9), 2610-2626.

data. In the best-case scenario, all participants with missing outcomes in the intervention group will be assumed to have the most favourable outcome, and all missing participants in the control group the least favourable outcome. In the worst-case scenario, the opposite assumptions will be applied. The robustness of the findings will be assessed by examining whether the direction and interpretation of the estimated treatment effects remain consistent across these extreme assumptions.

Compliance

As the ITT approach captures the averaged effect of offering Summer Jobs, treatment effects in the presence of compliance at the individual level will be looked at to capture the average effect of participation in Summer Jobs. For this analysis, young people will be classified as either compliant or non-compliant. To be compliant young people will need:

- To complete ≥80% of scheduled hours or content in preparation week; and
- To complete ≥60% of scheduled hours or content at placement.

Data on compliance will be provided by LDPs via the CRM.

We will further extend our analysis by using an Instrumental Variable (IV) approach to give an indication of the effects a) of receiving Summer Jobs and b) of the intensity of placement. To implement a) and b), we will use a Two Stage Least Squares (2SLS) approach with group allocation as the instrumental variable for compliance. In a), we will define compliance as a binary indicator that takes the value 1 if the young person is compliant and 0 otherwise. In b), we will define compliance as a continuous variable measuring the proportion of Summer Jobs scheduled hours that the young person received out of the total planned for participants. The first stage will model the compliance variable using the same explanatory variables used for the headline ITT analyses and the instrument. The second stage model will use predicted compliance in place of the group identifier variable in the ITT analyses specified above to generate Complier Average Causal Effect (CACE) estimates.

Most participants will receive some level of Summer Jobs, so variation in “dose” is limited. However, this level of participation is considered core to achieving any effect, so the CACE analysis will still provide insight into the impact of higher vs lower levels of engagement, rather than a strict “none vs full” comparison.

Presentation of outcomes

Effect sizes will be reported for all primary and secondary outcomes.

Binary outcomes (co-primary outcomes and EET) will be expressed as risk ratios (RRs) with natural frequencies.

Continuous outcomes (SDQ, JET aspiration, NGSC) will be expressed as Cohen's D, calculated as the difference in group means divided by the control-group standard deviation, with small-sample correction. For all continuous outcomes, histograms of observed scores will be presented to show the distribution of values within intervention and control groups, highlighting implications for interpreting effect sizes and suitability of measures.

Model-based standard errors will be used for all analyses, consistent with the specification of the primary, secondary, subgroup and compliance analysis models.

All p-values will be reported as continuous probabilities without dichotomisation around thresholds such as 0.05 and without the use of asterisks to denote significance. This applies to primary, secondary, subgroup, and compliance analyses.

Statistical uncertainty will be reflected using 95% confidence intervals (CIs) for all effect sizes. RRs will be derived from logistic regression models; Cohen's D will be calculated from group means and control-group standard deviations. Subgroup, compliance, and follow-up analyses will use the same metrics.

Outcome	Type	Effect Size Metric	Timepoint	Source
Offending	Binary	RR	Baseline	DfE–PNC Datashare
EET (co-primary)	Binary	RR	Baseline	LEO dataset
SDQ total difficulties	Continuous	Cohen's D	12–16 weeks	Self-report
EET (secondary)	Binary	RR	12–16 weeks	Self-report
JET aspiration to work	Continuous	Cohen's D	12–16 weeks	Self-report
NGSC total score	Continuous	Cohen's D	12–16 weeks	Self-report

Appendix A

Power Calcluations

```
# Offending
compute_mde_negative <- function(p1, n_total, power = 0.80, alpha = 0.03, attrition = 0) {

 n_eff <- n_total * (1 - attrition)
 n_per_group <- n_eff / 2

 find_diff <- function(diff) {
  p2 <- p1 - diff
  h <- 2 * asin(sqrt(p1)) - 2 * asin(sqrt(p2))   # Cohen's h
  achieved_power <- pwr.2p.test(h = h, n = n_per_group, sig.level = alpha)$power
  return(achieved_power - power)
 }

 mde <- uniroot(find_diff, c(0.001, p1 - 0.001))$root
 p2 <- p1 - mde
 RR <- p2 / p1  # Risk ratio
 h <- 2 * asin(sqrt(p1)) - 2 * asin(sqrt(p2))
 p_comb <- (p1 + p2)/2
 d_approx <- (p1 - p2)/sqrt(p_comb*(1 - p_comb))

 data.frame(
  N_total = round(n_total),
  Attrition = paste0(round(attrition*100), "%"),
  MDE_pct = round(mde*100, 2),
  Risk_Ratio = round(RR, 3),
  Cohens_h = round(h, 3),
  Cohens_d_approx = round(d_approx, 3)
 )
```

```
}

# EET
compute_mde_positive <- function(p1, n_total, power = 0.80, alpha = 0.03, attrition = 0) {

  n_eff <- n_total * (1 - attrition)
  n_per_group <- n_eff / 2

  find_diff <- function(diff) {
    p2 <- p1 + diff
    h <- 2 * asin(sqrt(p2)) - 2 * asin(sqrt(p1))   # Cohen's h positive
    achieved_power <- pwr.2p.test(h = h, n = n_per_group, sig.level = alpha)$power
    return(achieved_power - power)
  }

  mde <- uniroot(find_diff, c(0.001, 1 - p1 - 0.001))$root
  p2 <- p1 + mde
  RR <- p2 / p1  # Risk ratio
  h <- 2 * asin(sqrt(p2)) - 2 * asin(sqrt(p1))
  p_comb <- (p1 + p2)/2
  d_approx <- (p2 - p1)/sqrt(p_comb*(1 - p_comb))

  data.frame(
    N_total = round(n_total),
    Attrition = paste0(round(attrition*100), "%"),
    MDE_pct = round(mde*100, 2),
    Risk_Ratio = round(RR, 3),
    Cohens_h = round(h, 3),
    Cohens_d_approx = round(d_approx, 3)
  )
}

# Inputs
year2_offending <- 782
year3_offending <- 1650
year2_eet <- 747
year3_eet <- 1580

attritions <- c(0, 0.10, 0.20)

# Control rates
p_controls_offending <- c(0.30, 0.40)  # For offending (negative outcome)
p_controls_eet <- c(0.50, 0.65, 0.75)  # For EET (positive outcome)

# ---- Offending calculations ----
results_offending <- do.call(rbind, lapply(p_controls_offending, function(p1) {
```

```
  n_total <- year2_offending + year3_offending
  do.call(rbind, lapply(attritions, function(a) {
    cbind(ControlRate = p1, Outcome = "Offending", compute_mde_negative(p1, n_total,
attrition = a))
  }))
}))

# ---- EET calculations ----
results_eet <- do.call(rbind, lapply(p_controls_eet, function(p1) {
  n_total <- year2_eet + year3_eet
  do.call(rbind, lapply(attritions, function(a) {
    cbind(ControlRate = p1, Outcome = "EET", compute_mde_positive(p1, n_total, attrition =
a))
  }))
}))
```

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413