STATISTICAL ANALYSIS PLAN

Dallaglio RugbyWorks sports for development programme to reduce the risk of violence and offending: a two-armed randomised controlled trial

ICF Consulting Services Ltd

Principal investigator: Dr Matt Barnard

YOUTH ENDOWMENT FUND

YEF statistical analysis plan

Project title	Dallaglio RugbyWorks sports for development programme to reduce the risk of violence and offending: a two-arm randomised controlled trial
Developer (Institution)	Dallaglio RugbyWorks
Evaluator (Institution)	ICF Consulting Services Ltd
Principal investigator(s)	Dr Matt Barnard
SAP author(s)	Robert Wishart & Sergio Sallis
Trial design	Two-arm randomised controlled trial with random allocation at the individual level
Trial type	Efficacy
Evaluation setting	Pupil Referral Units, Alternative Provision schools, and mainstream schools with onsite exclusion provision
Target group	CYP aged 11 to 16 years old who have been permanently excluded from mainstream schools and/or have had three or more fixed term exclusions and are at risk of being involved in youth violence, gang activity, and/or crime as a perpetrator or victim
Number of participants	50 settings, 900 pupils
Primary outcome and data source	Behavioural difficulties, measured by the SDQ externalising score (survey data collection)

Secondary outcome and data source	Wellbeing, measured by the Warwick Edinburgh Mental Wellbeing Scale (WEMWBS, survey data collection) Pro-social behaviour, measured by the SDQ pro-social scale (survey data collection) Emotional difficulties, measured by the SDQ internalising score (survey data collection) Fixed-term exclusions (school management information) Attendance (school management information)

SAP version history

Version	Date	Changes made and reason for revision
1.0 [*original*]		*[leave blank for the original version]*

Table of contents

Introduction

This study aims to evaluate Dallaglio RugbyWorks' (DRW's) sports for development programme through a two-armed randomised controlled trial (RCT), randomised at an individual level and with the control group receiving business-as-usual from pupil referral units, alternative provision (AP) schools, mainstream schools with alternative provision, and in some cases children's social care services. The Dallaglio RugbyWorks programme aims to reduce behavioural difficulties through raising the aspirations of CYP, developing their life skills, focussing on their mental health and improving their physical wellbeing. Improvements in these outcomes are expected to lead to increased likelihood of the CYP being in education, employment and training (including reduced school exclusions) and reduced likelihood of violence and offending (impacts which are likely to be self-reinforcing). The study incorporates an implementation and process evaluation, which involves collecting quantitative and qualitative data from CYP in both intervention and control groups and from coaches and management staff. In addition, the study will include an analysis of the costs of delivery. The rationale behind the approach is that RCTs are an effective way of assessing the net impact of an intervention and an individual randomisation approach was selected as the most feasible and efficient way of generating a sufficient sample size.

The target population for the intervention is young people aged 11-16 exhibiting a 'secondary' level of need (CYP at high risk of becoming involved in violent or non-violent crime, based on the YEF eligibility triangle categorisation). This may include young people exposed to sexual exploitation. Training is provided to staff to be able to appropriately engage these young people – including understanding risk factors, appropriate interventions, trauma-informed approaches, and raising safeguarding concerns – and ensure these young people are not excluded from our programme. DRW staff do not offer therapy, however the programme can support young people to be able engage in other forms of support.

In terms of eligibility, any young person who is attending an AP school or a PRU will have been excluded from school and is therefore considered to be at high risk of becoming involved in violent or non-violent crime and is eligible to take part in the intervention. The nature of the structure of PRUs means that young people won't be 'referred' to the intervention, instead the whole cohort will be invited to take part in the trial.

In addition to recruiting from PRUs, CYP excluded in mainstream (EiM) will be recruited. These are CYP who are managed out with of the normal mainstream school experience in the following ways:

- in a separate building on the mainstream school site;

- in a suite of classrooms within the mainstream school;
- in the same classrooms pupils in “normal mainstream education” use but on separate timetables

Wherever they are based, all CYP invited to take part in the trial will meet the criteria below:

- They will be permanently excluded from mainstream and/or have had three or more fixed term exclusions.
- They will have been identified by the provision as displaying challenging behaviour (this will usually be evidenced through the provision’s behaviour management system).
- They will be at risk of offending due factors such as: or be living in a neighbourhood with high levels of crime and socio-economic deprivation.
- Have experience of trauma or higher number of adverse childhood experiences;
    - Have siblings, or be associated with peers, who are known to be involved with offending.
    - Have previously exhibited criminal behaviour or had involvement in youth justice services.

In addition, for settings to be eligible for the trial, a minimum of 12 eligible CYP need to be recruited. This ensures that sample sizes at analysis stage (after attrition) are big enough to support robust analysis.

Design overview

Trial design, including number of arms		Two-arm randomised controlled trial
Unit of randomisation		Individual child or young person (CYP)
Stratification variables (if applicable)		Pupil Referral Unit (PRU), Alternative Provision schools, and mainstream schools with onsite exclusion provision. Academic year of randomisation.
	variable	Behavioural difficulties

Primary outcome	measure (instrument, scale, source)	Externalising score based on the sum of the conduct and hyperactivity sub-scales of the SDQ (self-report version), 0 – 20, survey data collection.
Secondary outcome(s)	variable(s)	Wellbeing Pro-social behaviour Emotional difficulties Fixed term exclusions Attendance
	measure(s) (instrument, scale, source)	Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), 14-70 (survey data collection) Pro-social scale of the SDQ (self-report version), 0-10 (survey data collection) Internalising score based on the sum of two sub-scales (emotional difficulties and peer difficulties scales) of the SDQ (self-report version), 0–20 (survey data collection) Number of fixed term exclusions in one academic year, count (school management information) Percentage of all sessions missed in the academic year they were randomised due to authorised or unauthorised absence, 0-100 (school management information)
Baseline for primary outcome	**variable**	Behavioural difficulties
	measure (instrument, scale, source)	Externalising score based on the sum of two sub-scales of the SDQ (self-report version), 0–20 (survey data collection)
Baseline for secondary outcome	**variable**	Wellbeing Pro-social behaviour Emotional difficulties
	measure (instrument, scale, source)	Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), 14 - 70, survey data collection Pro-social scale of the SDQ (self-report version), 0-10, survey data collection

		Internalising score based on the sum of the conduct and hyperactivity sub-scales of the SDQ (self-report version), 0 – 20, survey data collection Percentage of all sessions missed in the academic year prior to randomisation due to authorised or unauthorised absence, 0-100 (school management information)

Sample size calculations overview

		Protocol	Randomisation
Minimum Detectable Effect Size (MDES)		0.186	
Pre-test/ post-test correlations[1]	level 1 (participant)	0.45	0.45
Alpha		0.05	0.05
Power		0.80	0.80
One-sided or two-sided?		Two-tailed	Two-tailed
Number of participants	intervention	360	
	control	360	
	total	720	

The sample size calculations are based on achieving a minimum detectable effect size (MDES) of 0.186, in line with YEF guidance and with average effect sizes for sports programmes on externalising behaviour reported in the literature (Gaffney, et al. (2021) converted using the

[1] Whilst the template asks for the pre-post correlation, the actual parameter relevant to the power calculations is the variance as explained by the baseline of the outcome and the blocking covariates (R^2); we assume this to be 0.20. As R^2 and pre-post correlation are different statistical concepts, a direct translation is not possible. We use the square root of the R^2 as an approximation of the pre-post correlation (0.45).

formula suggested by Chinn (Chinn, 2000)). The power calculations were conducted in PowerUp! (Dong & Maynard, 2013) for a two-level, fixed effects, blocked-individual random assignment design (2.2 BIRA2_1f). The power calculations were based on the following assumptions.

- A type-one (false positive) error rate of 0.05.
- A type-two (false negative) error rate of 0.20 (synonymous with power of 0.80).
- Two-tailed statistical significance testing.
- A 1:1 allocation ratio between intervention and control.
- The variance in the outcome expected to be explained by the baseline measure and blocking covariates is 0.20. We estimate this to be roughly equivalent to a pre-post correlation of 0.45.[1] This is a reasonable assumption given that the same outcome measure is being used at baseline and endline, twelve months apart. This correlation is also similar to assumptions from other trials using the same outcome, such as the Mentoring MST (Lewis, 2023), which assumes a pre-post correlation of 0.50.

The sample size assumptions at protocol stage were determined based on achieving the specified MDES and assumed that delivery would span two academic years. Specifically:

- Approximately 25 settings would be recruited for participation in the trial.
- On average, 18 young people will be recruited to the trial within each setting, each year.
- The total sample size would therefore consist of 900 young people (450 intervention; 450 control).
- The total sample size at analysis has been adjusted to account for attrition between baseline and endline; we assume 20% attrition. This assumes an analysis sample of 720 young people; 360 intervention, 360 control).

At the time of drafting version 1.0 of this SAP, the year one randomisation has been conducted. Year two randomisation will take place in Autumn 2025. In year one, 496 CYP were randomised (247 intervention; 249 control).

The trial was pre-registered with the ISRCTN registry (ISRCTN17394989)

Analysis

The analyses outlined in this section have been developed based on the YEF statistical analysis guidance (Youth Endowment Fund, 2021). This analysis plan has been drafted after the year 1 baseline data collection and randomisation, but before any endline data has been collected.

All analysis will be conducted in the latest available version of Stata[2]. Analysis code will be retained, and the primary analysis code will be published in an appendix of the final evaluation report.

Primary outcome analysis

The primary outcome will be **behavioural difficulties** as measured by the raw scores of the externalising behaviours score,[3] based on the self-report SDQ (Goodman, 1997).

The SDQ measures emotional and behavioural difficulties and has been widely used in clinical and research settings. It has good psychometric properties and is comprised of five sub-scales, which are: emotional symptoms, conduct problems, hyperactivity/inattention, peer problems, and pro-social behaviour. Each subscale has an individual score. The externalising behaviour score (which ranges from 0-20, where a higher score indicates more negative externalising behaviours) is the sum of the conduct and hyperactivity subscales. These will be derived using the scoring syntax available from the SDQ website. The primary outcome aligns with the intervention's intermediate outcomes, which are to improve mental and physical wellbeing, leading to reduced stress and better emotional self-regulation (which is reflected in a number of items that make up the conduct problems and hyperactivity subscales.

The primary outcome analysis will be conducted on an intention-to-treat (ITT) basis and will use a linear regression model, where the SDQ externalising behaviours raw score at endline is the dependent variable, regressed against the SDQ externalising behaviours raw score at baseline, a binary indicator of their allocation (intervention or control) and fixed effects for the block within which they were randomised (setting) and the academic year of randomisation.

The decision to use fixed effects to account for the blocking is informed by the YEF's statistical analysis guidance (Youth Endowment Fund, 2021) which recommends that fixed effects are more appropriate for efficacy trials where we do not attempt to generalise beyond the sites within the trial.

The regression model will use ordinary least squares (OLS). The model equation is outlined in Equation 1.

Equation 1 Primary analysis model

$$EBE_{ij} = \beta_0 + \beta_1 EBB_{ij} + \beta_2 Allocation_{ij} + \beta_3 Setting_j + \beta_4 Year + u_j + e_{ij}$$

[2] At the time of writing, this is StataNow 18.5.

[3] We propose raw scores for consistency with other YEF evaluations.

Where:

- EBE_{ij} is the externalising behaviours raw score at endline for individual *i*, in setting *j*.
- β_0 is the coefficient for the regression intercept.
- β_1 is the coefficient for the externalising behaviours raw score at baseline, EBB_{ij}.
- β_2 is the coefficient of interest; the outcome change for CYP allocated to the intervention group ($Allocation_{ij}$).
- β_3 represents the vector of coefficients for the vector of setting fixed effects, $Setting_j$.
- β_4 represents the coefficient for the academic year of randomisation (used in stratification).
- u_j and e_{ij} are the error terms for setting and individual levels respectively.

The primary analysis will apply the following Stata syntax:

reg post_eb pre_eb i.intervention i.setting i.academicyear

Secondary outcome analysis

There will be five secondary outcomes, collected for all pupils in intervention and control groups. Note that we do not formally adjust for multiple-hypothesis testing. All the secondary analysis results should be interpreted considering that there is a risk of type-one error inflation.

The first secondary outcome is **mental wellbeing** as measured by the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS). The WEMWBS was developed to enable the monitoring of mental wellbeing in the general population and the evaluation of projects, programmes and policies which aim to improve mental wellbeing. The WEMWBS is a 14-item scale of positively worded statements covering feeling and functioning aspects of mental wellbeing. The 14-statements have five response categories from 'none of the time' (scored one) to 'all of the time' (scored five). This means that the scale can take a value ranging from 14 to 70. CYP are asked to describe their experiences over the previous two weeks.

Mental wellbeing is identified in the theory of change as a key intermediate outcome and is directly relevant to the understanding the causal mechanism of the programme. Improved mental wellbeing is likely to reflect a reduced amount of stress and frustration experienced by the CYP. The frustration displacement hypothesis indicates that this will reduce their aggressive behaviour (a key component of behavioural difficulties) (King, 2012).

The second secondary outcome is **pro-social behaviour**, as measured by the raw scores of the pro-social scale on the SDQ. Pro-social behaviour is another important intermediate outcome of the intervention, reflecting its emphasis on raising aspirations leading to changes in CYP's goal values and goal expectancies (Morgenroth et al, 2015).

The third secondary outcome is **emotional difficulties,** as measured by the raw scores of the internalising behaviour scale on the SDQ.

The fourth and fifth secondary outcomes are indicators of engagement in **education, employment or training (EET)**, specifically **fixed-term exclusions** and **attendance** (both based on school management information). Alongside reductions in violence and offending, engagement in EET is a key impact of the intervention resulting from reduced behavioural difficulties. Engagement in EET is also likely to reinforce reduced behavioural difficulties and thus reduce the risk of CYP being violent or offending. The exclusion measure will be the total number of fixed-term exclusion (a count variable) and the absence rate is measured as the proportion of all possible sessions missed by a young person due to authorised and unauthorised absence. Both these measures will be collected for a single academic year; 2024-25 for the first randomisation cohort, and 2025-26 for the second randomisation cohort.

Mental wellbeing, pro-social behaviour, emotional difficulties and attendance will be measured using continuous variables for the purposes of regression analysis. Therefore, they will be analysed using OLS, in line with the primary analysis. They will use the same specification as the primary analysis, substituting the baseline and endline outcomes respectively. For each of these outcomes we will explore the distribution of scores and residuals to ensure the chosen regression model is appropriate. If necessary, we will employ a more appropriate regression model. For example, if an outcome illustrates a count distribution, we will use a Poisson regression model. The distributions and choice of regression model will be reported transparently. This will also apply to the primary analysis, if necessary.

The number of fixed term exclusions is a count variable, and we expect it to follow a Poisson distribution. Therefore, it will be analysed using a Poisson regression but will otherwise share the same specification as the other secondary outcome analyses.

Subgroup analyses

The study has not been powered for subgroup analysis, and therefore is considered exploratory (Youth Endowment Fund, 2021). The subgroup analyses were pre-specified in the trial protocol and will be conducted solely for the primary outcome, behavioural difficulties. There are three individual characteristics that define the subgroups which are of interest for this trial. These are ethnicity, gender and Special Educational Needs and Disabilities (SEND) status.

Based on the target group of beneficiaries, we anticipate that there will be a sufficient number of pupils from Black, Asian, or other minority backgrounds to conduct subgroup analysis by ethnicity. It is our intent to separately analyse different ethnicity groups against a consistent reference group of White pupils. This is preferable to a dichotomous analysis of White pupils

against pupils from all ethnicity groups combined, as this may hide impact heterogeneity. However, the feasibility of this subgroup analysis depends upon the numbers of pupils within each ethnicity group.

Error! Reference source not found. outlines the characteristics of the year one randomised sample by allocation. This implies that sub-group analysis by separate ethnicity groups is likely to be feasible. Equation 2 outlines the model equation for the sub-group analysis by ethnicity in this scenario. The model equation will therefore be:

Equation 2 Subgroup analysis by ethnicity

$$EBE_{ij} = \beta_0 + \beta_1 EBB_{ij} + \beta_2 Allocation_{ij} + \beta_3 Ethnicty_{ij} + \beta_4 EthniciyInteraction_{ij} + \beta_5 Setting_j + u_j + e_{ij}$$

Where $Ethnicty_{ij}$ represents the vector of ethnicity dummy variables and $EthniciyInteraction_{ij}$ represents the vector of interaction variables between random allocation and ethnicity dummy variables.

Table 1 Year 1 subgroup analysis characteristics by allocation

	Control	Intervention	Total
Ethnicity			
Asian or Asian British	12	19	31
Black, Black British, Caribbean or African	46	41	87
Mixed or Multiple ethnic groups	36	33	69
White	138	132	270
Other ethnic group	8	9	17
Missing, prefer not to say	8	12	20
Gender identity			
Male	184	190	374
Female	65	53	118
Other, prefer not to say	0	5	5
Special Educational Needs and Disabilities (SEND) status			
No health condition	159	161	320
Health condition	49	52	101
Missing, prefer not to say	41	34	75

The baseline survey asks two questions related to sex and gender. The first question asks for the respondent's sex (to which they can respond female, male or other/prefer not to say) and the second question asks if they identify with the same sex as they were registered at birth to which they can respond yes or no, with an option to enter their gender identity. Whilst sex refers to gender assigned at birth, it is possible that some young people may respond other/prefer not to say. We will use both questions to derive a variable for gender.

We anticipate that there will be too few young people answering "other/prefer not to say" or indicating an alternative gender identity to male/female for robust subgroup analysis of three or more subgroups. This is reinforced by the year one randomisation sample characteristics (see **Error! Reference source not found.**). Therefore, the subgroup analysis by gender will be dichotomous, using the responses to both questions to analyse the results by their current gender identity as male or female. Individuals who have responded with another gender identity will not be considered for this subgroup analysis. Similarly, those who have not responded will be excluded (complete case analysis). Therefore, the subgroup analysis by gender will have the following model equation.

Equation 3 Subgroup analysis by gender identity

$$EBE_{ij} = \beta_0 + \beta_1 EBB_{ij} + \beta_2 Allocation_{ij} + \beta_3 Gender_{ij} + \beta_4 GenderInteraction_{ij} + \beta_5 Setting_j + u_j + e_{ij}$$

Where $Gender_{ij}$ is a binary indicator that the individual identifies as female, and $GenderInteraction_{ij}$ is the interaction between random allocation and female gender identity.

Finally, the baseline survey asks two questions related to health and disability. Specifically, it asks "Do you have any physical or mental health conditions or illnesses lasting or expecting to last 12 months or more?" to which they can respond "Yes", "No", or "Prefer not to say". If they answer yes, they are also asked "Do any of your conditions or illnesses reduce your ability to carry out day-to-day activities?" to which they can respond "Yes, a lot", "Yes, a little", "Not at all", or "Prefer not to say". As we expect very few respondents to respond yes to the first question, it is unlikely that a more granular sub-group analysis will be possible. Therefore, we will conduct sub-group analysis based on the binary response to the first question – those who respond "prefer not to say" will be excluded from the analysis (complete case analysis). Therefore, the model will have the following equation.

Equation 4 Subgroup analysis for health condition/disability

$$EBE_{ij} = \beta_0 + \beta_1 EBB_{ij} + \beta_2 Allocation_{ij} + \beta_3 Health_{ij} + \beta_4 HealthInteraction_{ij} + \beta_5 Setting_j + u_j + e_{ij}$$

Further analyses

We will also conduct the following exploratory analyses:

- Analysing the total difficulties score of the SDQ and the individual sub-scales of the SDQ (emotional problems, conduct problems, hyperactivity, peer problems).
- Assessing treatment effect heterogeneity by estimating treatment effect by site.

The analyses of the total difficulties score of the SDQ and the individual sub-scales will follow the same approach as the primary analysis, substituting the relevant outcome at baseline and endline into the model.

The analysis of treatment effect heterogeneity will explore the distribution of effect sizes between settings. No settings will be identified (they will be reported anonymously as Setting 1, Setting 2, etc.). The effect sizes by setting will be estimated using a multi-level regression model where CYP (level one) are clustered within sites (level 2). The sites will be interacted with allocation to estimate the impact estimates by site. The model equation will be:

$$EBE_{ij} = \beta_0 + \beta_1 EBB_{ij} + \beta_2 Allocation_{ij} + \beta_3 Setting_j + u_{0j} + u_2 Allocation_{ij} + e_{ij}$$

Where $u_2 Allocation_{ij}$ represents the random slopes of for random allocation. The effect sizes and their respective confidence intervals will be reported graphically, showing the distribution of effect sizes from smallest to largest. We will also conduct a likelihood ratio test to compare this model with a model that does not include random slopes to assess whether the site-by-treatment interaction provides a better fit for the data.

Imbalance at baseline

Imbalance at baseline between intervention and control group CYP on observable characteristics will be reported on descriptively. As the randomisation is at individual level, setting-level characteristics will not be assessed. The descriptive statistics will include the following characteristics:

- Outcome variables: The baseline of the primary outcome (behavioural difficulties).
- Demographic characteristics: This will include ethnicity, gender and health/disability (in their raw form as asked in the questionnaire) as well as any other relevant characteristics that may be available.

Continuous variables will be reported as means and standard deviations, with statistical significance testing for differences between the intervention and comparison group using a t-test (or its non-parametric equivalent). Count variables (i.e., number of fixed term exclusions) will be re-coded as a categorical variables for the baseline balance tests (e.g. no exclusions/one/two/three/four/five or more). These categorical variables will be used to implement tests between the proportions observed for the intervention and control group (statistically significant differences between intervention and control groups will be judged using chi-square tests). These statistics will be reported as randomised and as analysed with p-values of less than 0.05 considered an indication of statistically significant differences between groups.

Missing data

Reasons for missing data will be fully explored in the participant flow diagram, which will identify the reasons for lossess and exclusions from recruitment to analysis. Missing data analysis will be conducted for the primary outcome if greater than 5% of the randomisation sample are missing outcome data that excludes them from the primary analysis (which uses the complete case sample).

Firstly, missing data will be explored descriptively by comparing the characteristics of individuals with no missing data to individuals who are missing data that precludes them from the primary analysis. The characteristics included will include all the individual level characteristics outlined in the Imbalance at Baseline section, as well as their random allocation and setting-level characteristics such as setting type (whether mainstream school, PRU or alternative provision school). Statistically significant differences will be explored in the same way as the Imbalance at Baseline analysis, using appropriate statistical significance testing (t-tests and Chi-square tests).

Additionally, a "drop-out" model will be estimated using logistic regression. The dependent variable will be a binary indicator of whether the individual has missing data that precludes them from being included in the primary analysis. The dependent variables will be all the variables outlined for the descriptive analysis above, where missing on these covariates is captured by an additional missing data category. If there are statistically significant associations between these characteristics and the dependent variable, we will assume that the data is Missing at Random (MAR).

Multiple imputation will be conducted if we assume data is MAR, and the proportion of missing data is greater than 5% of the randomised sample, but less than 30% of the randomised sample. The imputation will use all variables used in the "drop-out" model, with the variables included in their "raw" form. In addition, if the missing outcome data is conditional on covariates then a sensitivity analysis of the primary analysis will be estimated including this covariates in the regression model. The results of the MI and/or expanded regression model will be compared for consistency with the primary analysis results and interpreted in line with Figure 1 of the YEF analysis guidance (Youth Endowment Fund, 2021).

Multiple imputation will be conducted using chained equations (MICE) using *mi impute chained* in Stata. We will set a "random seed" to ensure the imputation can be replicated. We will discard the first 20 iterations ("burn-in") to allow for convergence of the model. At this stage, we assume that we will impute 100 datasets, though appropriate diagnostic statistics will be estimated to ensure the imputation is running correctly. If necessary, we will adjust our imputation approach (e.g. increasing the "burn-in" period or the number of imputed datasets).

Once the imputed datasets have been created, we will then re-analyse the primary analysis model as a sensitivity analysis using the imputed data. We will use the *mi estimate* command in Stata.

Compliance

It is difficult to define compliance for the Dallaglio intervention in a way that could be logically interpreted as part of a Complier Average Causal Effect (CACE) estimate. Compliance is likely to be one-sided, because the coaches control access to the intervention, so those allocated to the control group will not be allowed to receive the intervention. Amongst those in the intervention group, compliance could be considered to consist primarily of their attendance to sessions, along with ensuring that sessions are delivered as intended by coaches. However, with no quantitative, independent way of monitoring session delivery, the only information we can draw on is attendance to sessions.

If the average attendance to sessions in the intervention group is less than 75% of the total possible sessions, then a **complier average causal effect (CACE)** analysis will be conducted for the primary outcome. The CACE will be estimated using a two-stage least squares (2SLS) instrumental variables (IV) regression, using randomisation as the instrument (Angrist & Imbens, 1995).

Intra-cluster correlations (ICCs)

Intra-cluster correlation coefficients (ICCs) will be estimated for the primary outcome using a multilevel regression. The dependent variable is the primary outcome (SDQ externalising behaviours score at endline) regressed against the baseline SDQ externalising behaviours score, a binary indicator of random allocation and a random intercept for settings. A second ICC estimate will be provided for an empty model (i.e. primary outcome with no dependent variables and a random intercept for setting).

The ICC will be estimated using the following formula:

$$\rho_s = \frac{\sigma_{BS}^2}{\sigma_{BS}^2 + \sigma_{WS}^2}$$

Where:

- σ_{BS}^2 is the between setting variance; and,
- σ_{WS}^2 is the within setting variance.

Presentation of outcomes

The impact estimates will be presented as Hedge's g effect sizes, in line with YEF statistical analysis guidance (Youth Endowment Fund, 2021). This will be calculated as:

$$g = J \times \left(\frac{\bar{Y}^T_{adj} - \bar{Y}^C_{adj}}{SD_{pooled}} \right)$$

Where:

- $\bar{Y}^T_{adj} - \bar{Y}^C_{adj}$ is the adjusted mean difference in outcomes, given by the coefficient β_2 in the primary analysis model.
- SD is the pooled standard deviation, given by:

$$SD_{pooled} = \sqrt{\frac{(n_T - 1)SD_T^2 + (n_C - 1)SD_C^2}{n_1 + n_2 - 2}}$$

- J is the correction factor, given by:

$$J = 1 - (\frac{3}{4(n_T + n_C - 2) - 1}$$

Confidence intervals will be estimated as:

$$g - zv_g \leq g \leq g_{WT} + zv_g$$

Where z is the critical value from the normal distribution, multiplied by the standard error.

References

Angrist, J. D. & Imbens, G. W., 1995. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association,* 90(430), pp. 431-442.

Chinn, S., 2000. A simple mehtod for converting an odds ratio to effect size for use in meta-analysis. *Stat Med,* 19(22), pp. 3127-31.

Dong, N. & Maynard, R., 2013. PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quais-Experimental Design Studies. *Journal of Research on Educatonal Effectiveness,* 6(1), pp. 24-67.

Gaffney, H., Jolliffe, D. & White, H., 2021. *Sports Programmes: Toolkit Technical report,* London: YEF.

King, B. 2012. Psychological Theories of Violence. *Journal of Human Behaviour in the Social Environment*, 22:553-571

Goodman, R., 1997. The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry,* Volume 38, pp. 581-586.

Lewis, J., 2023. *Statistical Analysis Plan: IMpacts of a short-term mentoring model for young people: a multi-site randomised controlled trial,* London: YEF.

Youth Endowment Fund, 2021. *Resources for Evaluators; Statistical Analys Guidance.* [Online] Available at: https://youthendowmentfund.org.uk/resources-for-evaluators/ [Accessed 17 09 2024].

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413