STATISTICAL ANALYSIS PLAN

Impacts of a short-term mentoring model for young people: a multi-site randomised controlled trial

Centre for Evidence and Implementation (CEI), YMCA George Williams College (GWC) and Bryson Purdon Social Research (BPSR)

Principal investigator: Jane Lewis

YOUTH ENDOWMENT FUND

Impacts of a short-term mentoring model for young people: a multi-site randomised controlled trial
Statistical analysis plan

Evaluating institution: Centre for Evidence and Implementation, YMCA George Williams College, Bryson Purdon Social Research
Principal investigator(s): Jane Lewis

Project title[1]	**Impacts of a short-term mentoring model for young people: a multi-site randomised controlled trial**
Developer (Institution)	Centre for Evidence and Implementation (CEI), YMCA George Williams College (GWC), and Bryson Purdon Social Research (BPSR) with seventeen youth agencies.[2]
Evaluator (Institution)	Centre for Evidence and Implementation (CEI), YMCA George Williams College (GWC), and Bryson Purdon Social Research (BPSR)
Principal investigator(s)	Jane Lewis
SAP author(s)	Dr Susan Purdon, BPSR
Trial design	Two-armed randomised waitlist controlled trial with random allocation at the individual level
Trial type	Efficacy trial
Evaluation setting	Seventeen youth agencies delivering mentoring services in varied community contexts across the UK

[2] The Centre for Youth Impact (initially named partner on the project) merged with YMCA George Williams College and became a 'centre of expertise' within the College.

Target group	Primarily targeting 10-14-year-olds (with some 15-17-year-olds) with risk factors related to youth violence
Number of participants	17 youth organisations, N = 850 young people
Primary outcome and data source	Strengths and Difficulties Questionnaire (young person self-report online survey)
Secondary outcome and data source	Selected Evaluation of National Citizen Service domains and individual items (young person self-report online survey)

SAP version history

Version	Date	Changes made and reason for revision
1.2 [*latest*]		
1.1		
1.0 [*original*]	09/10/2023	*[leave blank for the original version]*

Contents

Introduction

This efficacy trial forms part of a wider study of multisite trials which is testing the feasibility of undertaking randomised controlled trials (RCT) of mentoring across multiple youth service sites.

The study of multisite trials consists of two parts: a feasibility trial completed in November 2022 and the efficacy trial described in this document. Seventeen youth agencies (referred to here as 'delivery partner organisations' or DPOs) have been recruited to take part. The small-scale feasibility RCT was successfully delivered by nine of these DPOs with the report on that study due to be published later in 2023. The second phase of this study was initially framed as a pilot trial, to explore the feasibility of delivering a shared model of practice and running an RCT with a larger number of organisations and participants. However, due to the success of the feasibility trial and the planned recruitment numbers, it was decided that the trial was well placed to explore impact. In light of this, the study has been re-formulated as an efficacy trial.

The study aims to generate important learning about how to undertake multi-site trials with youth organisations and of non-manualised practices. The primary question addressed by the efficacy trial will be: What is the impact of short-term mentoring on the social and emotional learning skills of young people at risk of youth violence, compared with services as usual?

Other aims of the efficacy trial are to undertake an implementation and process evaluation detailed in the evaluation protocol. [3]

The intervention is based on a shared practice model of mentoring, with this model being developed as part of the feasibility stage. The aim was to develop a model of mentoring practice sufficiently consistent for a trial but flexible enough to align with DPOs' usual practice and not obstruct the objective of evaluating mainstream, non-manualised youth provision.

The core programme structure for the shared practice model includes:

- a minimum of 12 weeks duration
- 12 sessions of at least 45 minutes over the course of 12 weeks

[3] The protocol for the efficacy trial can be found on the YEF website for more details - https://youthendowmentfund.org.uk/funding/who-we-fund/multi-site-trial-mentoring/

- adult (rather than peer) mentors who are paid (rather than volunteers)
- mentoring delivered on a one-to-one basis
- voluntary participation of the mentee
- mentees predominantly aged 10-14, with no more than 30% aged 15-17
- mentees, at the beginning of the study, exhibiting at least one of the YEF-listed 'unmet needs' from the YEF Outcomes Framework[4]

Those in the control group will receive 'services as usual' i.e., the typical provision provided by the DPO or by an agency to which they refer a young person, including group sessions, sports, and trips, but excluding one-to-one support. The intervention group would similarly be expected to receive services as usual as well as mentoring.

The Strengths and Difficulties Questionnaire (SDQ) total difficulties score will provide the primary social and emotional learning outcome measure. The secondary social and emotional learning outcomes are belief-based measures from the Evaluation of National Citizen Service, focusing on: (a) self-confidence (leadership and communication); (b) problem-solving and decision-making skills; (c) teamwork and social skills building; (d) resilience/emotional regulation.

Design overview

The efficacy trial is being run as an RCT with 17 DPOs and with two arms per DPO: an intervention arm, and a wait-list control arm[5]. The allocation is in the ratio 50:50 per DPO.[6] Each DPO aimed to recruit a minimum of 50 young people, with a total sample of approximately 850. Outcomes data will be collected at baseline and 12-week follow-up.

DPOs are responsible for recruiting young people to the trial. For those recruited and deemed eligible, consent is collected from both parent/carer and young person, and the young person completes an online baseline questionnaire including the primary and secondary outcome measures prior to randomisation. The DPOs will complete basic demographic details (including age, ethnicity, gender and whether this is the same as sex assigned at birth) about

[4] YEF's Outcomes Framework https://youthendowmentfund.org.uk/outcomes/ detailing the factors in a young person's life that can influence their likelihood of becoming involved in crime and violence.

[5] The feasibility trial looked at the perceived impacts of the mentoring through interviews with young people, mentors, and DPO managers. No unintended negative consequences were highlighted, so we expect the control group to be offered mentoring irrespective of the findings on impact

the young person. Once the consent has been collected and the baseline questionnaire completed, the young person can enter the trial, with random allocation taking place at that point. The randomisation algorithm is based on a merged block randomisation procedure[7] which is appropriate for multi-site trials.

Those randomised to the intervention arm start mentoring as soon after randomisation as is feasible; those allocated to the control arm are eligible to start mentoring after 12 weeks, once they have completed their follow-up outcomes survey.

Data on outcomes is collected at two points in time per young person: baseline and at, or soon after, 12 weeks post randomisation. Supporting data is being collected throughout the trial, covering:

- demographics (baseline)
- number of sessions of mentoring (intervention arm only)
- receipt of other services during the trial (both arms)
- whether any control group member started mentoring before the 12-week outcome survey was completed

An online 'data portal' has been developed to capture all of the quantitative trial data. This includes an embedded randomisation tool.

Table 1: Trial design overview

Trial design, including number of arms		Two-arm randomised waitlist controlled trial
Unit of randomisation		Individual young person
Stratification variables (if applicable)		DPO
	variable	Social and emotional learning skills[8]

[7] Merged block randomisation: A novel randomisation procedure for small clinical trials. Stephanie L van der Pas. Clinical Trials (2019) Vol 16(3) 246-252

[8] The SDQ does not, strictly speaking, directly measure SEL. However, the key outcome that mentoring is expected to change in the short-term is SEL, and the SDQ was chosen as the closest available validated measure to this. See the study protocol for a fuller discussion.

Primary outcome	measure (instrument, scale, source)	Strengths and Difficulties Questionnaire total difficulties score, young people self-report (Goodman et al, 1998), fielded in online survey 12 weeks after randomisation.
Secondary outcome(s)	variable(s)	Self-confidence; problem-solving/decision-making; teamwork/social skills building; emotional regulation/resilience
	measure(s) (instrument, scale, source)	Self-report items from the Evaluation of the National Citizen Service (Fitzpatrick et al, 2021), fielded in online survey 12 weeks after randomisation
Baseline for primary outcome	**variable**	Social and emotional learning skills
	measure (instrument, scale, source)	Strengths and Difficulties Questionnaire total difficulties score, young people self-report (Goodman et al, 1998), fielded in online survey prior to randomisation
Baseline for secondary outcome	**variable**	Self-confidence; problem-solving/decision-making; teamwork/social skills building; emotional regulation/resilience
	measure (instrument, scale, source)	Self-report items from the Evaluation of the National Citizen Service (Fitzpatrick et al, 2021), fielded in online survey prior to randomisation

Sample size calculations overview

Each DPO was set a target of recruiting and randomising 50 young people and delivering mentoring to them, with 50 being set as a challenging, but achievable, number. The expectation was that, across all the DPOs, this would give a trial of around 850 young people, around 425 per arm. In practice some DPOs found they could not meet the target of 50 and the actual number recruited and randomised was somewhat lower at 744.

The target sample size for the trial was established at the point when the trial was being planned as a pilot rather than as an efficacy trial. DPOs were recruited with the understanding that they would recruit at least 50 young people each, this being considered a good, yet challenging, test of whether small DPOs could manage trials of this size. Nevertheless, with 17 DPOs, this would give a large overall sample size of at least 850 young people recruited.

For a trial of the planned size it was estimated that effect sizes of around 0.17 standard deviations would be detectable with 80% power. This is line with the effect sizes found in other trials of mentoring where effects sizes that average at 0.21sd have been found across a range of studies and outcomes[9]. The MDES was calculated within Excel using the formula[10]:

$$MDES = (1.96 + 0.84)\sqrt{\frac{2}{n}(1 - R^2)}$$

where n is the achieved sample size per arm (=425) and R is the correlation between baseline and follow-up outcomes. For the primary SDQ outcome, R was assumed to be 0.5, this being the value found in the feasibility trial (unpublished statistic). The value 1.96 is the z-value for a type I error rate (alpha) of 0.05, and 0.84 is the z-value for 80% power (type II error rate of 20%).

The MDES for the actual achieved randomisation sample size is slightly larger at 0.18 standard deviations. If there is attrition per arm of around 10%, giving an analysis sample per arm of 335, an effect size of 0.19 standard deviations will be detectable.

Table 2: Sample size calculations

		Protocol	Randomisation
Minimum Detectable Effect Size (MDES)		0.17sd	0.18sd
Pre-test/ post-test correlations	level 1 (participant)	0.5	0.5
	level 2 (cluster)	0	0
	level 1 (participant)	0	0

[9] Raposa, E.R., Rhodes, J., Stams, J.M., Card, N., Burton, S., Schwartz, S., Yoviene Sykes, L.A., Kanchewa, S., Kupersmidt, J. and Hussain, S. (2019) *The Effects of Youth Mentoring Programs: A Meta-analysis of Outcome Studies Journal of Youth and Adolescence*, 48:423-443

[10] See for example Section 7.1.2 of Djimeu, E.W., and Houndolo, D-G. (2016) Power calculation for causal inference in social sciences. International Initiative for Impact Evaluation Working Paper 26.

		Protocol	Randomisation
Intracluster correlations (ICCs)	level 2 (cluster)	0	0
Alpha[11]		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two	Two
Number of participants	intervention	425	372
	control	425	372
	total	850	744

Analysis

The analysis of the efficacy trial data will be on an intention-to-treat basis. Estimates of impact per outcome will be regression-based, with the baseline version of each outcome being entered as a covariate. DPO will be entered as a fixed effect. The analysis will be conducted in SPSS v28.0.1.1.

Primary outcome analysis

The primary outcome measure for the efficacy trial, will be the Strengths and Difficulties Questionnaire (SDQ)[12], a core measure of social and emotional learning included in most YEF evaluations.

The SDQ is a validated scale with an established evidence base which measures behaviours, emotions, and relationships across 25 items. The efficacy trial has adopted the self-report

[11] Please adjust as necessary for trials with multiple primary outcomes, 3-arm trials etc. when a Bonferroni correction is used to account for family-wise errors.

[12] Goodman, R (2001) Psychometric properties of the strengths and difficulties questionnaire. Am Acad Ch Adolesc Psychiatry 40 (11) 1337-45.

version, suitable for 11- to 17-year-olds.[13] It includes five subscales, each with five items, that measure: 1. Emotional symptoms; 2. Conduct problems; 3. Hyperactivity/inattention; 4. Peer problems; 5. Prosocial behaviour. Young people score from 0 to 2 on each item using a scale 'not true', 'somewhat true' or 'certainly true', thus producing a score for each subscale from 0 to 10. The primary outcome in the analysis of the efficacy trial will be the total 'difficulties' score (from 0 to 40), calculated by summing the first four subscales. Exploratory analysis will include the three subscales: the prosocial subscale (score 0 to 10), 'internalising problems' (which combines emotional and peer symptoms) and 'externalising problems' (which combines conduct and hyperactivity symptoms), with the latter two scores ranging from 0 to 20. These three subscales have been tested and used effectively by the scale developers in low-risk and general population samples (Goodman et al, 2010 and are known to provide intermediate risk and protective factors of offending If we find a significant effect of the intervention on the total score, the exploratory analysis will help determine what is driving that.

The calculation of the scores will follow the standard SDQ scoring rules. The total difficulties score will only be calculated where all four subscales have a valid score (that is, at least three of the five items have been answered), others being set to missing. Likewise, the internalising and externalising scales will only be calculated where both subscales within the scale have a valid score. Each subscale will be calculated as (total subscale score)*5/(number answered).

The main regression model specification is as follows:

$$(SDQF)_{i,j} = \alpha + \beta_1 Group_{i,j} + \beta_2 (SDQB)_{i,j} + \beta_3 DPO_j + \varepsilon_i \qquad \text{(Eq 1)}$$

where i=young person belonging to DPO=j; SDQF is the SDQ score at follow-up; SDQB is the SDQ score at baseline; Group is set equal to 1 if young person within a DPO belonged to the intervention group and 0 otherwise, DPO_j represent the organisation level dummy variables capturing the DPO level fixed effects, and ε_i is the individual level error.

As an exploratory analysis, a second model will be run to test whether there is a significant difference in impacts across the organisations in the trial. This will be tested by including an interaction term to the above regression model ($Group_{i,j} * DPO_j$). If we find evidence of between DPO differences we will report the impact estimate for each DPO graphically, but not undertake pair-wise tests.

Secondary outcome analysis

[13] Parent/carer and teacher versions are also available.

Four secondary outcome measures will measure shorter-term social and emotional learning outcomes, with the questions being taken from previous evaluations of the National Citizen Service (NCS).[14] These items have been chosen as likely to reflect changes in outcomes after 12 weeks of mentoring. More discussion on the rationale for selecting these is included in the study protocol.[15]

The decision to use the NCS items is based on (a) the face validity of the items, which speak directly to the Theory of Change, and (b) whilst not validated, their proven sensitivity in other studies to change over a three-month period. Overall, 21 NCS items are included which, between them, cover the following belief-based domains: (a) self-confidence: leadership and communication (seven items); (b) problem-solving and decision-making skills (four items); (c) teamwork and social skills building (six items); (d) resilience/emotional regulation (four items). Seven items use a five-point confidence scale, from 'very confident' to 'not at all confident', while the others use a five-point Likert scale ('strongly agree' to 'strongly disagree'). Although the NCS evaluations have reported separately on the impact on each item, we will use principal axis factor analysis to produce four separate outcomes, one per domain.

For the factor analysis per domain, each item within the domain will be scored from 1 to 5, with 5 being the most positive response. A single factor will be generated per domain. Those with missing data on any of the items within a domain will be treated as missing for the domain.

The analysis of each domain will be conducted following the same model specification as the primary SDQ outcome.

Subgroup analyses

Exploratory analysis will be undertaken to establish whether there is evidence of sub-group differences in the efficacy of mentoring by:

- gender (male; female);
- ethnic group (white; mixed; Asian/Asian British; Black/Black British; other).

[14] The latest published report is Fitzpatrick et al (2021): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1015222/NCS_2019_Evaluation_Report.pdf

[15] https://youthendowmentfund.org.uk/wp-content/uploads/2023/03/Mentoring-MST-Protocol-report-March-23.pdf

This will be a descriptive analysis without formal testing of differences (for which the trial is underpowered). The effect sizes per sub-group will be generated via a regression model with interaction terms (randomisation group by sub-group) added. This exploratory work will be presented alongside any qualitative research findings on sub-group differences.

Further exploratory analyses

It is possible that we conclude from the main analysis that there is between-organisation variation in effectiveness. If we do identify variation, we will undertake exploratory regression analysis to gain an understanding of what is driving those differences. Given the findings are likely to be quite tentative, we will look to triangulate them with the IPE. For this analysis DPOs will be included as a random, rather than fixed effect in the main regression model. These analyses will consider the following factors:

- location of provision (i.e., school-based versus other)
- the profile of young people (in terms of demographics and baseline scores)
- attrition rates
- compliance amongst the intervention group.

As stated earlier, there will be exploratory analysis looking at the effect sizes for three SDQ subscales: the prosocial subscale, 'internalising problems', and 'externalising problems'. Further exploratory sub-group analyses will also be undertaken as already described.

Imbalance at baseline

The trial report will summarise baseline characteristics per randomisation group for all young people randomised, for those with a valid total SDQ score at both baseline and follow-up (i.e. the primary analysis subset), and for those with valid baseline and follow-up scores for each of the four domains in the secondary outcomes. The differences at baseline will provide evidence on whether the randomisation gave balanced groups; the differences at follow-up will provide evidence on whether attrition has introduced an imbalance.

The characteristics shown will be: number of young people randomised per DPO, gender, age, ethnic group, SEND status, SDQ total difficulties score mean at baseline, , the four secondary outcome means at baseline. Where average scores on primary and secondary outcomes are reported, standard deviations will be reported.

Missing data

We are anticipating two types of missing data: unit missings, where young people fail to complete a follow-up questionnaire, and item missings where young people do not answer all of the questions put to them. The item missings may occur at both baseline and follow-up.

We will describe and summarise the extent of missing data per outcome, and the reasons for data being missing will be reported on. We will document how much missing data is unit non-response and how much is item non-response, and whether there are associations between data being missing and the baseline outcomes and young person characteristics.

The primary ITT regression analysis will be based on complete cases, that is those for which all of the variables needed for the model are complete. This assumes that missing data is 'missing completely at random' (MCAR). However, if missing data means that we need to exclude more than 5%[16] of cases in a model then we will assess the sensitivity of the results to alternative assumptions about the mechanisms leading to missing data.

Firstly, if for any outcome the percentage of cases excluded from the analysis because of missing data is at least 5%, we will use the baseline outcome variables and young person characteristics to model (via a logistic regression) the probability per outcome of a case being excluded from the analysis because of missing data. The regression model (Eq 1) will be re-run to include, as covariates, the predictors of exclusions that are identified from this logistic regression. This will help establish whether the effect sizes are influenced by the level and nature of missing data, under an assumption of missing at random (MAR).

In addition, if a regression model covariate has more than 5% missing, then multiple imputation (MI) will be used to generate multiple data sets with imputed covariates. We will re-estimate the effect sizes per outcome variable across these imputed datasets, taking the average of the effect sizes generated as the best estimate of the effect size under MAR assumptions

Finally, if there is evidence that outcomes data is missing not at random (MNAR) we will include some estimates of effect sizes based on a range of extreme assumptions about the missing outcomes. This will generate upper and lower bounds for the effect sizes. (MNAR might occur if, for example, outcomes data was not able to be collected from YP in the intervention group who concluded that mentoring was not working well for them and who dropped out of the trial, or from YP in the control group who concluded they no longer needed mentoring and dropped out of the trial). The assumptions adopted for this sensitivity analysis will include imputing the worst possible outcomes scores for those missing from the intervention arm and the best possible outcomes scores for those missing from the control arm.

Compliance

[16] This is in line with YEF guidance. If less than 5% of cases are missing in an analysis it is unlikely that the effect sizes derived from the model would be biased.

There is no formal definition of compliance. The feasibility trial suggests most young people allocated to the intervention group will receive at least some mentoring sessions. In the feasibility trial, 35 out of 46 in the intervention arm completed the target number of eight mentoring sessions, and a further five completed at least six. All attended some sessions, the minimum being two. Just one out of 47 in the control group received any mentoring during the evaluation period. The target number of sessions has been increased to 12 for the efficacy trial, and a smaller percentage may meet this, but we still anticipate the great majority of those in the intervention group will attend multiple sessions.

Given this anticipated high rate of compliance, we do not plan to conduct a complier average causal effect (CACE) analysis[17].

Presentation of outcomes

Effect sizes will be calculated using Hedges' g, as specified in the following equation:

$$ES = \frac{(\overline{Y_T} - \overline{Y_c})_{adjusted}}{s^2}$$

where $\overline{Y_T}$ is the regression adjusted mean for the treatment group, $\overline{Y_c}$ is the regression adjusted mean for the control group (computed using Eq 1), and s^2 is the pooled unconditional variance of the two groups.

Effect sizes will be reported along with confidence intervals and p-values to reflect statistical uncertainty.

[17] A related reason is that we do not expect to be able to generate unbiased CACE estimates. For unbiased estimation the number of sessions attended would need to be strongly associated with the pre-programme YP characteristics. Our expectation is that a lot of the partial compliance will be attributable to other, unrelated, factors, such as staff absences rather than being related to the characteristics of the YP.

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413