STATISTICAL ANALYSIS PLAN

Media Academy Cymru's (MAC) Cerridwen Project. A randomised control trial efficacy study with internal pilot.

Cordis Bright

Principal investigators: Matt Irani, Suzie Clements, Dr Stephen Boxford, Prof Darrick Jolliffe, Kam Kaur, Madeleine Morrison, Ashna Devaprasad

YOUTH ENDOWMENT FUND

Media Academy Cymru's Cerridwen Project

Statistical analysis plan

Evaluating institution: Cordis Bright

Principal investigator(s): Matt Irani, Suzie Clements, Dr Stephen Boxford, Prof Darrick Jolliffe, Kam Kaur, Madeleine Morrison, Ashna Devaprasad

Project title	Media Academy Cymru's (MAC) Cerridwen Project. A randomised control trial efficacy study with internal pilot.
Developer (Institution)	Media Academy Cymru
Evaluator (Institution)	Cordis Bright
Principal investigator(s)	Matt Irani, Suzie Clements, Dr Stephen Boxford, Prof Darrick Jolliffe, Kam Kaur, Madeleine Morrison, Ashna Devaprasad
SAP author(s)	Prof Darrick Jolliffe, Suzie Clements, Dr Stephen Boxford, Matt Irani, Madeleine Morrison, Ashna Devaprasad
Trial design	Two-armed parallel randomised control trial with random allocation at the young person level
Trial type	Efficacy trial with internal pilot and implementation and process evaluation
Evaluation setting	The intervention is delivered at a variety of safe spaces in the young people's community
Target group	10- to 17-year-olds in Cardiff, Swansea, Merthyr Tydfil and Caerphilly who are at risk of involvement in serious violence or exploitation

Number of participants	596 (298 in treatment group, 298 in control group)
Primary outcome and data source	Self-reported offending (volume score on the Self-Reported Delinquency Scale) (See, Smith & McVie, 2003)
Secondary outcome and data source	Empathy (Basic Empathy Scale) (Jolliffe and Farrington, 2006) Pro-social values and behaviour (SDQ – Pro-social behaviour subscale) (Goodman, 2005) Behavioural difficulties (SDQ – externalising behaviours score (combining conduct problems and hyperactivity/inattention subscales)) (Goodman, 2005)

1 SAP version history

Version	Date	Changes made and reason for revision
1.2 [*latest*]		
1.1	Revised May 2025	Changed SDQ scale from conduct problems subscale to externalising behaviours score (combining conduct problems and hyperactivity/inattention subscales) *Reason: We agreed to change our approach during the internal pilot* Updated delivery area to include Caerphilly

		Reason: We, MAC and YEF agreed an expansion of the delivery area as part of the internal pilot Added further modelling into sample size calculations *Reason: To reflect learning from the pilot period*
1.0 [*original*]	Published March 2025	*[leave blank for the original version]*

2 Table of contents

3 Introduction

This is the statistical analysis plan (SAP) for an efficacy study including a two-armed parallel randomised controlled trial (RCT) and implementation and process evaluation of Media Academy Cymru's (MAC) Cerridwen programme. This plan should be read in conjunction with the study trial protocol, which is available at: https://youthendowmentfund.org.uk/wp-content/uploads/2024/07/MAC-Cerridwen-Evaluation-Protocol.pdf [1]

The Cerridwen programme is a six-month voluntary, one-to-one mentoring and case management intervention, rooted in cognitive behavioural approaches. The programme is working with young people living in the areas of Cardiff, Swansea, Merthyr Tydfil and Caerphilly in South Wales between April 2024 and April 2026.

Through a youth work approach, Cerridwen aims to reduce children and young people's future engagement in violence and offending behaviour by enhancing young people's empathy, building their emotional resilience, improving wellbeing, and challenging negative narratives. For a full list of the programme's intended short-term, medium-term, and long-term outcomes, please see the trial protocol.

The rationale for an efficacy RCT of Cerridwen is strong; in the UK there is limited robust evidence for what works to reduce offending among children and young people, and more research is needed for mentoring programmes with a specific focus on children and young people already involved in crime or violence. An efficacy RCT of Cerridwen will therefore contribute to knowledge and understanding of what works to reduce offending for this cohort. More information on the study rationale and background is available in the trial protocol.

The efficacy RCT which included an internal pilot (more about this can be viewed in the study trial protocol) started in April 2024, and will run until July 2026, with final reporting taking place in December 2026.

The evaluation team is collecting a range of data throughout the trial, including:

- Self-reported outcomes measures, including the Self-Reported Delinquency Scale (volume score subscale), the Social Support and Rejection Scale (full measure), the Basic Empathy Scale (full measure), and the Strengths and Difficulties Questionnaire (pro-social behaviour and externalising behaviours subscales).

[1] Last accessed 7 January 2025.

- Demographic data provided by MAC, including sex, age, race/ethnicity, involvement with other services, special educational needs and disabilities (SEND) data, and English as an Additional Language (EAL) status.
- Activity and monitoring data collected by the MAC delivery team.

This document sets out our planned analysis for each data type in more detail. The rest of this document is structured in the following way:

Section 4: Trial design sets out the research questions, key outcomes and measures, and randomisation approach for the trial.

Section 5: Sample size calculations presents the power calculations for the trial.

Section 6: Analysis sets out the approach to analysis of primary outcomes and secondary outcomes, subgroup analysis, exploratory analysis, further analysis, missing data, compliance and outcomes presentation.

4 Trial design

4.1 Overview

The efficacy study is a two-armed parallel RCT, comparing the effectiveness of a dedicated mentoring and case management programme (treatment group) to light-touch young person-led wellbeing and safety support (control group) in reducing children and young people's future engagement in youth violence and offending behaviours. All young people living in the areas of Cardiff, Swansea, Merthyr Tydfil and Caerphilly who are referred into the project, who are 10-17, who meet the eligibility criteria and who consent to be part of the evaluation are randomly allocated to either the treatment or control group on a 1:1 basis.[2]

Recruitment is taking place on a rolling basis between April 2024 and September 2025. MAC have established various referral routes in partnership with a range of key referral organisations including statutory and non-statutory organisations. These include (but are not limited to):

- Social Services (Children's Services).
- Schools and Pupil Referral Units.
- Youth Services.

[2] Young people are eligible for Cerridwen if they meet all three of the following inclusion criteria: (1) Young people are exhibiting or are at risk of exhibiting violent behaviours, (2) Young people are living in the areas of Cardiff, Swansea, Merthyr Tydfil or Caerphilly, and (3) Young people are willing to engage with and complete Cerridwen. For more detail, please see the evaluation protocol.

- Youth Offending Services.
- Voluntary and Community Sector Organisations.[3]
- Self-referrals including young people and parents/carers.[4]

Referral sources and the number of eligible and ineligible referrals are monitored and will be reported in the final evaluation report.

Randomisation is conducted at the individual (young person) level, without stratification, in randomly varying block sizes of four, six and eight young people. The randomisation sequence was generated using an online randomisation service (Sealed Envelope).[5] Team members responsible for analysis were not involved in generating the randomisation sequence. For more information on the design of the randomisation process, please see the evaluation protocol.

Those recruited and randomly allocated to the treatment group receive weekly, two to three-hour one-to-one case management/mentoring sessions over a six-month period. Those randomly allocated to the control group receive light-touch, structured signposting and safeguarding support, provided by MAC. For more detail on the participant journey, please see the trial protocol.

Figure 1 below presents an overview of the trial design.

Figure 1: Trial design overview

Trial design, including number of arms		Two-arm parallel randomised control trial with random allocation at the young person level
Unit of randomisation		Individual young person
Stratification variables (if applicable)		None
	variable	Self-reported offending (violent or non-violent or general)

[3] Examples of organisations include: Atal Y Fro, Action for Children, Amber Project, Barnardo's, Fearless, Llamau, Platform, St Giles, The Hangout and Women's Aid.

[4] Self-referrals may form a small proportion of overall referrals. Self-referrals will be subject to the same assessment of eligibility as other referrals. Eligibility and consent are re-confirmed in the first meeting with the young person to ensure only the intended cohort access Cerridwen.

[5] See: https://www.sealedenvelope.com/. last accessed 10 January 2025.

Primary outcome	measure (instrument, scale, source)	Self-Reported Delinquency Scale (volume score)
Secondary outcome(s)	variable(s)	Empathy Pro-social values and behaviours Behavioural difficulties
	measure(s) (instrument, scale, source)	Empathy, measured by the Basic Empathy Scale (Jolliffe and Farrington, 2006). Pro-social values and behaviours, measured by the Strengths and Difficulties Questionnaire pro-social behaviour subscale (SDQ) (Goodman, 2005). Behavioural difficulties, measured by the Strength and Difficulties Questionnaire externalising behaviours scale (combining conduct problems and hyperactivity/inattention subscales) (SDQ) (Goodman, 2005).
Baseline for primary outcome	**variable**	Self-reported offending (violent and non-violent or general)
	measure (instrument, scale, source)	Self-Reported Delinquency Scale (volume score)
Baseline for secondary outcome	**variable**	Empathy Pro-social values and behaviours Behavioural difficulties
	measure (instrument, scale, source)	Empathy, measured by the Basic Empathy Scale. Pro-social values and behaviours, measured by the Strengths and Difficulties Questionnaire pro-social behaviour subscale (SDQ). Behavioural difficulties, measured by the Strength and Difficulties Questionnaire externalising behaviours scale (combining conduct problems and hyperactivity/inattention subscales) (SDQ).

4.2 Research questions

The primary research question of the impact evaluation is:

> *Is a dedicated mentoring and case management programme delivered with children and young people involved in (or at risk of involvement in) youth violence and offending behaviours, focused on understanding and managing emotions, an effective approach to reducing children and young people's future engagement in youth violence and offending behaviours compared to light-touch young person-led wellbeing and safety support?*

Secondary research questions are:

1. **Delivery:** Can the Cerridwen programme work under ideal circumstances?
2. **Impact:** a) What is the impact of the Cerridwen project? b) Do different subgroups of young people have different outcomes, e.g. those from minoritised/marginalised groups?
3. **Unintended consequences:** a) Does the Cerridwen project have any unintentional consequences? If so, what are these? b) Do different groups of young people experience these differently?
4. **Iatrogenic effects**: Are there any serious negative effects that can be attributed to the Cerridwen project on any outcomes?
5. **Mechanisms:** a) How does the Cerridwen project work to reduce young people's future engagement in offending? b) Which factors contribute most to the observed outcomes?

4.3 Outcomes

The primary outcome measure for the evaluation is self-reported offending as measured by the volume score on the Self-Reported Delinquency Scale (Smith & McVie, 2003) at baseline and T2, i.e., five months post-randomisation.[6]

About the SRDS Volume Score

[6] Please note T2 data is collected at five months post-randomisation as the last (sixth) month of Cerridwen is aimed at supporting young people to exit the programme safely.

The SRDS is a 19-item self-reported delinquency scale developed as part of the Edinburgh Study of Youth Transitions and Crime (Smith & McVie, 2003). It covers a range of antisocial and offending behaviours, has been validated for use with young people in the UK, and has been used with those aged between 10 and 17. The volume score is the sum of the number of times that a young person reports involvement in the 19 different items.

The SRDS has been shown to have good psychometric properties; reported internal consistency is between .87-.92 with an inter-item correlation of .19 (Fonagy et al., 2018; Humayun et al., 2017) and the measure correlates with official police arrests (89.5% - 95.2%; McAra & McVie, 2005).

More information on the subscales, psychometric properties and validity of the SRDS is available in the YEF outcomes measures database (Youth Endowment Fund, 2022b)[7] and in the YEF core measurement guidance (Youth Endowment Fund, 2021a and 2022a). [8]

The secondary outcomes that we are investigating are:

- Empathy, measured by the Basic Empathy Scale (BES) at baseline and T2 (five months post-randomisation).
- Pro-social values and behaviours, measured by the Strengths and Difficulties Questionnaire (SDQ) – pro-social behaviour subscale at baseline and T2 (five months post-randomisation).
- Behavioural difficulties, measured by the Strengths and Difficulties Questionnaire – externalising behaviours score at baseline and T2 (five months post-randomisation).

About the scales used to measure secondary outcomes

The **BES** is a 20-item self-report measure of empathy. It includes two subscales of affective and cognitive empathy. Each item is scored on a five-point Likert scale. A higher score reflects higher levels of empathy. The BES has convergent and divergent validity, and the Cronbach's alpha coefficients for the affective and cognitive empathy subscales is .79 and .85, respectively. More information on this scale, including psychometric properties and validity, is available in Jolliffe and Farrington (2006 and 2021).

[7] See: https://youthendowmentfund.org.uk/outcomes/. Last accessed 9 January 2025.

[8] See: https://res.cloudinary.com/yef/images/v1623145465/cdn/19.-YEF-SRDS-guidance/19.-YEF-SRDS-guidance.pdf and https://youthendowmentfund.org.uk/wp-content/uploads/2022/04/18.-YEF-SDQ-guidance-April-2022.pdf. Last accessed 9 January 2025.

The full **SDQ** is a 25-item questionnaire measuring behaviours, emotions and relationships for 4- to 17-year-olds. It contains five subscales:

1. Emotional symptoms.

2. Conduct problems.

3. Hyperactivity/inattention.

4. Peer problems.

5. Prosocial behaviour.

Each item is scored on a three-point Likert scale from 0 to 2, such that the scores for each subscale ranges from 0 to 10. For the prosocial values subscale high scores are desirable (i.e., greater prosocial values), but for the externalising behaviour scale (combining conduct problems and hyperactivity subscales) high scores are not desirable (i.e., a high score means greater externalising behaviours). The SDQ has been shown to have good internal consistency (Cronbach's alpha = .73), cross-informant correlation (mean=0.34), and retest stability after four to six months (mean: 0.62) (Goodman, 2001).

All primary and secondary measures were selected in agreement and collaboration with colleagues from MAC and YEF.

Data for all measures is collected directly from young people using either online survey software (SmartSurvey – see: https://www.smartsurvey.co.uk/) or a paper questionnaire at baseline and at five-months post-randomisation.

5 Sample size calculations

5.1 Overview

The primary objective of the Efficacy Study is to assess the effect of the Media Academy Cymru's (MAC) Cerridwen programme on self-reported offending as measured by the volume score on the Self-Reported Delinquency Scale (Smith & McVie, 2003). A sample size of 592 (296 per group), based on detecting a Minimum Detectable Effect Size (MDES) of 0.20 (equivalent to a 10% difference in proportions), assuming a correlation between baseline and follow-up of 0.5 (Humayun et al., 2019) and using a two-sided alpha of 0.05, would provide 80% power (Figure 2). Our assumptions about the MDES are informed by the YEF guidance (2021a) and previous research. For example, a meta-analysis using a random effects model (d=.21, 95% confidence interval, .07 to .34) of 18 studies, showed that mentoring programmes similar to Cerridwen make a 10-11% difference in relation to offending (Jolliffe & Farrington, 2008). We have also included the pre-post correlation based on values obtained from unpublished data from an RCT using the same outcome measure and in a similar population of adolescents (Humayun et al., 2019).

We have calculated the overall sample size for the trial a priori, in line with YEF guidance (2021a and 2021b). This is presented in the table below. The final sample for the trial will not be finalised until randomisation is complete in October 2025. We will update this table with the final figures once they are known.

These calculations were made with the Powerup! software package (Dong, N. and Maynard, R. A., 2013).

Figure 2: Sample size calculations

		Protocol	Randomisation
Minimum Detectable Effect Size (MDES)		0.20	
Pre-test/ post-test correlations	level 1 (participant)	0.5	
	level 2 (cluster)	N/A	
	level 1 (participant)	N/A	

		Protocol	Randomisation
Intracluster correlations (ICCs)	level 3 (cluster)	N/A	
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	
Number of participants	intervention	296	
	control	296	
	total	592	

The above table shows the Power Calculation for the study and suggests a total sample of 592 young people is required for a MDES of 0.20. However, we recognise that Cerridwen will need to receive more referrals than this to reach the target number due to anticipated attrition. We have worked closely with MAC colleagues taking into account YEF guidance to model target recruitment and retention rates needed to reach this sample of 592, factoring in an estimated attrition rate of approximately 10% from referral to recruitment, and a further 10% attrition from recruitment to completion of the intervention. This modelling suggests that the study needs to receive around 752 referrals to reach the final target sample size of 592. More information about modelled target recruitment and retention rates are available in the study protocol.

In the case of higher-than-expected attrition or lower-than-expected recruitment rates that result in the sample of 592 being exceeded or not reached, we have also calculated the possible MDES for different sample sizes in the table below. While the original modelling was based on estimated referral rates, updated recruitment and participation modelling informed by data from the internal pilot period suggests that a sample size of 367 is now more realistic. This would result in an MDES of 0.25. These estimates are based on the same approach, assumptions and software outlined above.

MDES	N recruited	Attrition %	N analysis
0.20	673	12	592
0.25	412	10	367
0.25	448	21	367
0.29	320	12	282
0.39	180	12	158
0.49	114	12	100
0.59	80	12	70

Based on: Pre-test/post-correlation of 0.50, p<.05, Power=.80, two-sided.

6 Analysis

6.1 Overview

This section presents the analytical approach for the Cerridwen efficacy study.[9]

All analysis will be conducted once delivery comes to an end in April 2025. Analysis will be conducted on an intent-to-treat basis in line with YEF guidance. This means that all participants will be analysed according to their allocation, regardless of whether they received the Cerridwen intervention or not. This provides the most conservative estimate of impact, as this approach evaluates the impact of offering the intervention to those who do and do not comply and helps to fully capture the 'true' benefit (if any) of the intervention (Torgerson and Torgerson, 2008). This approach is particularly relevant for policymakers and commissioners, i.e. those who may roll out an intervention but do not have control over take-up of the intervention across the system.

The analytical approach has been developed a priori and will be conducted in SPSS 25.[10]

6.2 Primary outcome analysis

The primary outcome is the volume of offending at the individual level as measured by the self-reported delinquency scale (SRDS; Smith & McVie, 2003), completed before randomisation and at five-months post randomisation. Young people are asked to report behaviours they have been involved in over the previous five months.

There is considerable debate about best practice when it comes to the analysis of data from RCTs. For example, Twisk et al. (2018) advocate for utilising longitudinal analysis of covariance or a repeated measures analysis without the treatment variable, but with the interaction between treatment and time in the model controlled for. They argue that failure to control for baseline differences in outcomes between the groups can lead to biased treatment estimates. Alternatively, others have cautioned against this approach (Sen, 2013).

Based on our understanding of such literature, our analysis of the impact of the Cerridwen programme on the SRDS volume score will be conducted using a fixed effects analysis of co-variance (ANCOVA) model, controlling for baseline SRDS score and site. The ANCOVA analysis will calculate the mean difference in SRDS scores between young people who received support from Cerridwen (the treatment group) and those who received 'light touch' wellbeing

[9] In line with YEF requirements this plan has been written before all baseline data for the efficacy study has been collected.

[10] IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

support (the control group), whilst accounting for baseline SRDS score and site. The outputs from this analysis will be used to calculate the Hedges' G effect estimate (described in Section 6.11). We have adjusted for site because Cerridwen, although managed by one central team, is being delivered in three areas (Cardiff, Merthyr Tydfil, Swansea and Caerphilly) by area-specific case managers who do not work across multiple sites. Adjusting for site will therefore help isolate and understand the influence of location-specific factors (such as geographic variation and staff influence) on the primary outcome.

We will use the following model:

$$Y_i = \beta_0 + \beta_1 Baseline_i + \beta_2 Group_i + \delta Area_k + \varepsilon_i \, \epsilon \, N(0, \sigma^2)$$

For $i = 1, \dots, n$ young people per area, and $k = 1, \dots, K$ Area

Where:

- Y_i is the T2 (five months post-randomisation) volume score for self-reported delinquency as measured by the SRDS.
- $Baseline_i$ is the baseline outcome measure of the SRDS volume score for young person i.
- $Group_i$ is a dummy variable for allocation group, i.e. 1 for the treatment group and 0 for the control group.
- β_2 is the average treatment effect, i.e. the primary parameter of interest for the trial.
- $Area_k$ is a vector of $K - 1$ binary area dummy variables.
- ε_i is the error/residual.
- σ^2 is the variance.

Specifying this model upfront will help to ensure that the analyses avoid the "fishing problem" and the "curse of dimensionality" (Humphreys et al 2013; Hayes, 2011).

We will also run robustness checks to assess the underlying assumptions for ANCOVA (e.g., linearity, homogeneity of regression slopes, normality of residuals, homoscedasticity) before conducting this analysis to ensure the proposed approach is appropriate. This will include assessing normality of the data using histograms and K-S tests. If the data does not meet these assumptions, we will run a non-parametric ANCOVA analysis. This will be determined once all data has been collected.

It is possible that the baseline and outcome variables may be skewed. Skew will be assessed using the traditional criteria based on their distribution (i.e., skews of greater or equal to 1.0

or less than or equal to -1.0). We will make a final decision about the best approach to address skew once this has been identified in the final data set. However, research suggests that the Tobit or two-part approach might be most appropriate (Boulton & Williford, 2018).

6.3 Secondary outcomes analysis

Our approach to analysis of secondary outcomes will mirror the approach outlined above for primary outcome analysis.

The secondary outcomes are:

- Ability to understand and/or experience the emotions of others (empathy), as measured by the Basic Empathy Scale (BES) (Jolliffe and Farrington, 2006) at baseline and T2 (five months post-randomisation).
- Pro-social values and behaviours measured by the Strengths and Difficulties Questionnaire pro-social behaviour subscale (SDQ) (Goodman, 2005) at baseline and T2 (five months post-randomisation).
- Behavioural difficulties, measured by the Strength and Difficulties Questionnaire externalising behaviours scale (SDQ) (Goodman, 2005) at baseline and T2 (five months post-randomisation).

For each secondary outcome we will conduct analyses of impact on the measures outlined above. This will be conducted using the same ANCOVA model specified for the primary outcomes measure outlined in Section 6.2 Primary outcome analysis above, which will include the baseline measurement for the respective secondary outcome variable where relevant.

We will also run robustness checks to assess the underlying assumptions for ANCOVA for each secondary outcome measure. This will include assessing normality of the data using histograms and K-S tests. If the data does not meet these assumptions, we will run a non-parametric ANCOVA analysis. This will be determined once all data has been collected.

These analyses will be conducted in SPSS 25.[11]

6.4 Subgroup analyses

The subgroup analyses we plan to undertake are likely to be exploratory in nature. Before undertaking any sub-group analyses, we would assess whether these would be sufficiently powered based on the data we have collected. If any subgroup analysis is not sufficiently powered, the analysis would be reported as exploratory, and any results caveated to be interpreted with caution.

We will assess the presence of heterogenous treatment effects in line with race equity, equality, diversity and inclusion considerations. There is limited evidence about the relative effectiveness of mentoring programmes for those from minority ethnic backgrounds. As such, we will explore whether it might be possible to evaluate whether the Cerridwen intervention was equally effective for those from racially minoritised backgrounds compared to those from White backgrounds. This would likely be an underpowered analysis so caution should be applied when interpreting the results.

We will conduct this ANCOVA analysis by exploring the presence of interaction effects between ethnicity and treatment allocation. We will report the estimated differences across subgroups with the respective confidence intervals. This would use the following model:

$$Y_i = \beta_0 + \beta_1 Baseline_i + \beta_2 Group_i + \gamma Area_k + \delta Ethnicity_{ij} + \theta Ethnicity_{ij} * Group_i + \varepsilon_i \in N(0, \sigma^2)$$

For $i = 1, \dots, n$ young people per area, with $j = 1, \dots, J$ ethnicities, in $k = 1, \dots, K$ areas

[11] IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

This model uses the same variables as the model set out in Section 6.2 Primary outcome analysis. In addition:

- $Ethnicity_{ij}$ is a vector of binary dummy variables for $J - 1$ ethnicities.
- θ is a vector of parameters indicating the existence of heterogenous treatment effects by ethnicity. The total treatment effect for young people in each ethnicity grouping will be $\beta_2 + \theta$.

We will report both the point estimates and confidence intervals for θ. If θ indicates that the Cerridwen intervention is differentially impactful for different ethnicities, we will also consider re-running the model in Section 6.2 Primary outcome analysis for each sub-group separately (i.e. for young people from White, Asian, Black, Mixed and Multiple ethnic backgrounds, and Other ethnic backgrounds). If the two treatment effects are similar, this will strengthen the findings from this exploratory analysis.

6.5 Further analyses

We will conduct the following exploratory subgroup analyses:

- **The impact of positive relationships.** We will evaluate the extent to which positive relationships between the young person and case manager (treatment group), or significant adult (control group) influenced the primary outcome over and above the impact of the Cerridwen Intervention (as measured by the SSRS – see call-out box below). This analysis is proposed because the Cerridwen theory of change suggests that a key mechanism of change for the intervention is that it has its effect through an increase in positive relationships with a trusted adult. This will take a mediation analysis approach, i.e. we will estimate the direct and indirect effects, following the approach outlined in Gunzler et al. (2013).

About the Social Support and Rejection Scale (SSRS)

The **SSRS** has four dimensions: Feels valued, trust, mentoring, and negativity. Each item is scored from 1 (never) to 5 (always). Each subscale score is the average of items that make up the subscale. Higher scores on the negativity scale reflect higher levels of stress and negativity within the relationship. For the overall scoring of the scale a high score represents a positive relationship.

- **Dosage and fidelity.** Fidelity to the model will be measured by the number and type of sessions participants receive, as well as how much time they are involved in Cerridwen. This will be compared to Cerridwen's theory of change and its intended delivery model. Any analysis relating to dosage and fidelity will be exploratory in nature. If power calculations suggest that this analysis would be sufficiently powered, we will explore

the associations between activities received as part of Cerridwen the level of dosage and the impact on the SRDS Volume score. This data will be captured by monitoring data on the number of sessions, and time received by Cerridwen participants collected by the Cerridwen delivery team. This analysis will be conducted using a general linear model assuming normality, or a generalized linear model. Possible variables will include the number of sessions received and the time period over which they were received. This approach would address questions such as: does attending eight or more Cerridwen sessions result in a similar impact as attending all sessions?

- **The impact on police contacts**. We will also explore the possibility of analysing reduced offending as measured by police data. We anticipate requesting data on police contacts, including type and timing of events (e.g., police contact, arrest, recorded offences, and convictions, all with associated dates). However, final data availability will be determined based on discussions with data providers. If we are able to access robust and reliable police data, we will explore whether Cerridwen had an impact on reducing contacts with police over and above that reported in the control group. This analysis would be conducted using a general linear model, repeated measures design (assuming normality) or a generalized linear model. We will include a treatment by outcome interaction term in the analysis. We will only proceed with this analysis if power calculations suggest these analyses would be sufficiently powered.

- **Exploration of unintended negative effects.** Through the analysis outlined in this SAP we will also explore whether the programme has any negative impacts on young people. We will triangulate this with data collected through the IPE.

6.6 Interim analyses and stopping rules

After the first cohort of baseline data collection we will analyse the completeness, reliability and validity of outcomes questionnaires (including the outcomes measures described above). We will do this by exploring:

- Percentages of scale item completeness.
- Outcome measure means, standard deviations and skew.
- Cronbach Alpha testing for scale reliability.

This analysis will not include a comparison between control group and treatment group data nor analysis of impact. We will review and discuss these findings with MAC and YEF to provide reassurance that data collection is proceeding well. If there are concerns, we will suggest and discuss solutions with YEF and MAC.

We will continue to monitor data quality looking at scale completeness, means, standard deviations and skew throughout the trial for internal purposes to ensure that data collection proceeds smoothly. If, on review, we have concerns, we will raise this with YEF and MAC as appropriate.

The trial will stop if the Cerridwen is unable to recruit a sufficient number of participants. Recruitment rates will be regularly monitored against modelled target rates and reviewed as part of project group meetings. In addition to recruitment concerns, attrition rates will also be monitored throughout the trial. Given the importance of minimising bias due to participant dropout, we will track:

- Overall attrition – the percentage of participants lost between baseline and endline.
- Differential attrition – differences in attrition rates between treatment and control groups.

Stopping decisions will be guided by YEF's evidence quality framework. If attrition exceeds 31%, we will assess the implications for statistical power and risk of bias with YEF and MAC colleagues. If attrition is high but manageable, we will consider mitigation strategies such as sensitivity analysis or imputation. However, if attrition reaches a level where the study can no longer produce meaningful findings, we will discuss stopping the trial with YEF and MAC colleagues. Any decisions about stopping will be made in discussion with YEF and MAC colleagues.

The Cerridwen project team will also be responsible for safeguarding of participants. They will report any serious adverse events overall and by trial arm. The trial will stop if MAC, YEF and Cordis Bright decide that the Cerridwen intervention is unsafe for participants.

6.7 Imbalance at baseline

If randomisation has been successful, both treatment and control groups should be equivalent at baseline. As such, any imbalance will have occurred by chance. To check for and monitor imbalance, we will produce a table of descriptive characteristics for all young people who have completed a baseline questionnaire. We will also produce an equivalent table for those who have completed a T2 questionnaire to check whether any attrition experienced throughout the trial may have introduced an imbalance. These descriptive characteristics will include age, sex, ethnicity (collected by Cerridwen colleagues) and the relevant outcomes collected through questionnaires (SRDS Volume Score, BES score, SDQ externalising behaviours and prosocial scale scores, SSRS score). Categorical data will be summarised by numbers and percentages. Continuous data will be summarised by mean (SD) if data are normal and median [IQR] if data are skewed.

Depending on the quality and completeness of data received, we may also include other data collected being collected by the Cerridwen team (i.e. involvement with other services, special educational needs and disabilities (SEND) data, and English as an Additional Language (EAL) status).

We will present a cross-tabulation of counts and percentages for each category above against allocation group. For continuous variables, we will present the means and standard deviation. This analysis will be used to inform our understanding of the extent to which our initial sample was balanced across the two groups, and whether any attrition experienced throughout the trial has introduced an imbalance. We will discuss any differences and their implications in the final report.

6.8 Missing data

Throughout the trial, the evaluation team will work closely with the Cerridwen intervention team to support the collection of high quality and complete data for all young people. However, missing data may arise due to either item non-completion or sample attrition (i.e. young people who do not complete all items of the baseline or the T2 questionnaires). We will assess both the extent of missingness and patterns of missingness in the data. In line with YEF guidance (2021a) we will report on both: (1) the proportion of missing data in the trial, and (2) the extent and pattern of missingness in the data. This will involve analysis of whether data is missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). MCAR and MAR mean that complete cases are unlikely to be biased subsequent to adjustment but may be underpowered, while MNAR suggests that structural bias has been introduced to the sample.

We will attempt to establish the missing mechanism (i.e. which variables in the data are predictive of non-response) through logistic regression models. This will model the presence of missing outcomes data with additional information that may be predictive of missingness. We will conduct this analysis in line with the flow chart in Figure 1 in the YEF analysis guidance (2021a). This outlines the following approach:

- If the prevalence of missing data is less than 5%, no further action is required as complete case analysis is unlikely to be biased.
- If outcomes data is MAR conditional on co-variates, we will include these co-variates in our primary analysis model and discuss the implications in full.
- If a covariate is MAR conditional on other covariates, we will conduct multiple imputation (MI). Treatment effects from the MI analysis will be reported in addition to estimates from the model outlined in Section 6.2 Primary outcome analysis. Any differences between the two and their implications will be discussed in full.

- If missing data cannot be fully explained by the other variables in the dataset, data is likely to be MNAR. In this scenario we will conduct a sensitivity analysis alongside the primary impact analyses.

We will only conduct the above analyses for the primary outcome analysis, as all secondary outcomes analysis, subgroup analysis and further analysis is tentative and exploratory in nature.

There is no universally agreed approach to analyses in the event of item non-completion. In the event that a high proportion of cases would be excluded due to low rates of item non-completion (for example, if most young people miss a small number of items), our approach to missing data will balance considerations around data integrity with maximising statistical power. In this scenario, we would consider using statistical techniques to impute missing items, in line with the YEF analysis guidance (2021a).

6.9 Compliance

As outlined in Section 6.1 Overview all analyses will be conducted on an intent-to-treat basis. This means that overall compliance for the purposes of the efficacy study will be met when young people have been randomised and allocated into the treatment or control group.

However, we acknowledge that intent-to-treat analysis may underestimate the efficacy of the intervention if some young people in either trial arm do not adhere to their assigned treatment. To examine this, we will conduct Complier Average Causal Effect (CACE) analysis, which will indicate treatment effects amongst those who comply with the intervention. However, any analysis of treatment effects in the presence of non-compliance will also be exploratory. This will be estimated using two-stage least squares (2SLS) regression (Gerber and Green, 2012), which uses the following two stages:

1. The first stage will model the compliance variable (i.e. number of sessions) using the same explanatory variables used for the primary analysis. This will be a logistic regression model used to generate predicted compliance.

2. The second stage models will use predicted compliance in place of the allocation group variable in the ITT primary analysis specified in Section 6.2 Primary outcome analysis to generate the CACE estimates.

We will report the results from the first stage of the 2SLS, along with the correlation between the instrument and endogenous variable and the associated F-test. Interpretations of the CACE estimates will be provided in the final report.

6.10 Intra-cluster correlations (ICCs)

This is not a clustered randomised controlled trial. As such, ICCs will not be calculated.

6.11 Presentation of outcomes

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

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{sd_{pooled}}$$

Where:

- $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the ANCOVA difference in means between the treatment and control groups adjusted for baseline outcomes measures and area, as specified in the primary outcomes model.

  sd_{pooled} is the unconditional pooled standard deviation of the two groups.[12]

With a sample of greater than 20 there is limited difference with Cohen's d. However, if the standard deviations between the treatment and control group are different, we would propose to use Glass' delta, which only uses the control group's standard deviation (Lipsey & Wilson, 2001).

We will report the statistical uncertainty associated with the effect sizes using both 95% confidence intervals (CIs) and two-tailed p-values, with statistical significance assessed at the conventional threshold of $p < 0.05$. Confidence intervals will be calculated using the following formula:

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

Where:

- Φ^{-1} is the percent point function of the normal distribution.
- g_{se} the standard error of the g statistic (noted as *ES* above).

All estimations and their statistical uncertainty will be reported, and the implications of both the point estimates and confidence intervals will be set out. In addition, all reporting will consider findings in light of the existing evidence base. This will be triangulated with the evidence collected from the implementation and process evaluation on the quality and context of delivery, the existence of theoretical causal mechanisms, and the experiences and

[12] $sd_{pooled} = \sqrt{\frac{(n_1-1)sd_1^2 + (n_2-1)sd_2^2}{n_1+n_2-2}}$, where n_1 and n_2 are the sample size for groups 1 and 2 respectively, and sd_1 and sd_2 are the standard deviations of group 1 and group 2 respectively.

perspectives of young people, parents and carers, key staff members, and wider stakeholders who participate in semi-structured interviews.

7 References

Akram, M., Cerin, E., Lamb, K. E., & White, S. R. (2023). Modelling count, bounded and skewed continuous outcomes in physical activity research: beyond linear regression models. *International Journal of Behavioral Nutrition and Physical Activity*, *20*(1), 1-11.

Boulton, A. J., & Williford, A. (2018). Analyzing skewed continuous outcomes with many zeros: A tutorial for social work and youth prevention science researchers. *Journal of the Society for Social Work and Research*, *9*(4), 721-740.

Dong, N. and Maynard, R. A. (2013). *PowerUp!*: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*,*6*(1), 24-67. doi: 10.1080/19345747.2012.673143

Fonagy P, Butler S, Cottrell D, Scott S, Pilling S, Eisler I, Fuggle P, Kraam A, Byford S, Wason J, Ellison R, Simes E, Ganguli P, Allison E, Goodyer IM. (2018). Multisystemic therapy versus management as usual in the treatment of adolescent antisocial behaviour (START): a pragmatic, randomised controlled, superiority trial. *Lancet Psychiatry* 5 (2), 119-133, doi: 10.1016/S2215-0366(18)30001-4.

Gaffney, H., Jolliffe, D. and White, H. (2022). *Mentoring Toolkit technical report* Mentoring-Technical-Report_Final.pdf (youthendowmentfund.org.uk) [Accessed 15/09/23].

Goodman, R. (2001). Psychometric Properties of the Strengths and Difficulties Questionnaire. *Child and Adolescent Psychiatry,* 40(11), 1337-1345.

Goodman, R., (2005). *The Strengths and Difficulties Questionnaire.* Available at: https://www.sdqinfo.org/a0.html [Accessed 17/8/23].

Gunzler, D., Chen, T., Wu, P. and Zhang, H., 2013. Introduction to mediation analysis with structural equation modeling. *Shanghai archives of psychiatry*, *25*(6), p.390.

Humayun, S., Herlitz, L., Chesnokov, M., Doolna, M., Landau, S., Scott S. (2017). Randomised controlled trial of Functional Family Therapy for offending and antisocial behaviour in UK Youth. *Journal of Child Psychology and Psychiatry* 58(9), 1023-1032.

Humphreys, M., De la Sierra, R.S. and Van der Windt, P., 2013. Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. Political Analysis, 21(1), pp.1-20.

Hayes, B., 2011. An adventure in the Nth dimension. American Scientist, 99(6), pp.442-446.

Jolliffe, D., & Farrington, D. P. (2006). *Basic Empathy Scale (BES)* [Database record]. APA PsycTests. https://doi.org/10.1037/t42339-000

Jolliffe, D. and Farrington, D.P. (2008). The influence of Mentoring on Reoffending *Swedish National Council for Crime Prevention*. Available at:

Jolliffe, D., & Farrington, D. P. (Eds.). (2021). Empathy versus offending, aggression and bullying: Advancing knowledge using the Basic Empathy Scale. Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780429287459

https://bra.se/download/18.cba82f7130f475a2f1800024229/1371914733834/2008_influence_of_mentoring_on_reoffending.pdf. [Accessed 15/09/23].

Lipsey, M. and Wilson, D. (2000) *Practical meta-analysis.* Thousand Oaks, CA: Sage.

McAra, L., McVie, S. (2005). The usual suspects? Street-life, young people and the police. *Criminal Justice* 5(1), 5-36.

Roffman, J. G., Pagano, M. E., and Hirsch, B. J., (2000). *Social support and rejection scale*. Evanston, IL: Human Development and Social Policy, Northwestern University.

Senn, S. (2013). Seven myths of randomisation in clinical trials. *Statistics in medicine*, *32*(9), 1439-1450.

Smith, D.J and McVie, S. (2003). Theory and method in the Edinburgh study of youth transitions and crime. *British Journal of Criminology* 43, 169–95.

Torgerson, D. J., & Torgerson, C. J. (2008). Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction. London: Palgrave Macmillan.

Twisk J., Bosman, L., Hoekstra T., Rijnhart, J., Welten, M. & Heymans M. (2018). Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications*, 10, 80-85.

Youth Endowment Fund, (2021a). *Analysis Guidance.* Available at: https://res.cloudinary.com/yef/images/v1623145483/cdn/6.-YEF-Analysis-Guidance/6.-YEF-Analysis-Guidance.pdf [Accessed 20 /10/23].

Youth Endowment Fund, (2021b). *Core measurement guidance: Self-Report Delinquency Scale.* Available at: https://res.cloudinary.com/yef/images/v1623145465/cdn/19.-YEF-SRDS-guidance/19.-YEF-SRDS-guidance.pdf. [Accessed 20 /10/23].

Youth Endowment Fund. (2022a) *Core Measurement Guidance: Strengths and Difficulties Questionnaire* https://youthendowmentfund.org.uk/wp-content/uploads/2022/04/18.-YEF-SDQ-guidance-April-2022.pd*f* Accessed 15/09/23.

Youth Endowment Fund. (2022b). *Outcomes Framework* https://youthendowmentfund.org.uk/wp-content/uploads/2022/08/YEF-Outcomes-Framework-August-2022.pdf Accessed 15/09/23.

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413