STATISTICAL ANALYSIS PLAN

Another Chance Fund Focused Deterrence programme: a multicentre randomised controlled trial

University of Hull; University College London; University of Abertay; University of Oxford

Principal investigator: Iain Brennan

YOUTH ENDOWMENT FUND

Another Chance Fund Focused Deterrence programme: a multicentre randomised controlled trial

Statistical analysis plan

Evaluating institution: University of Hull
Principal investigator(s): Iain R Brennan
SAP Authors: Alex Sutherland, Iain Brennan

YEF statistical analysis plan

Project title[1]	***Another Chance Fund Focused Deterrence programme: a multicentre randomised controlled trial***
Developer (Institution)	*Leicester, Leicestershire and Rutland Violence Reduction Network; Greater Manchester Combined Authority; West Midlands Violence Reduction Partnership; and Nottingham Violence Reduction Partnership*
Evaluator (Institution)	*University of Hull; University College London; University of Abertay; University of Oxford*
Principal investigator(s)	*Iain Brennan*
SAP author(s)	*Alex Sutherland, Iain Brennan*
Trial design	*Multicentre, two-arm, stratified randomised controlled trials with random allocation of individuals*
Trial type	*Effectiveness*
Evaluation setting	*Community*
Target group	*14+ year olds at risk of involvement in violence or already involved in violence.*
Number of participants	*Approximately 2,900 individuals*
Primary outcome and data source	*Count of violence against the person offences; police records (Police National Computer & local police data)*
Secondary outcome and data source	*Time to offence (police records); co-offending (police records).*

SAP version history

Version	Date	Changes made and reason for revision
1.1 [*latest*]	20/6/2025	See Changes from protocol table
1.0 [*original in protocol*]	24/02/2024	*[leave blank for the original version]*

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

Changes from protocol

Change made	Reason for revision
Amendment of outcome to 'charged, cautioned or convicted' of an offence. The previous definition of the primary outcome was an offence 'attributed to' an individual, which is imprecise and unstandardised as it could include an individual being a 'suspect' with little supporting evidence.	'Attributed to' is imprecise, can vary between police force areas and within forces over time, has a lower threshold and is susceptible to a range of individual-level biases and discrimination. The study is powered based on the anticipated distribution of the revised outcome.
Increasing timeline of baseline data from 12m to 24m for primary outcome and limiting to violence against the person offences.	To increase statistical power on the basis that this will provide more data on the study participants' prior offending which reduces between group variance and increases power through greater precision of estimates while also limiting to violence against the person offences, which aligns better with the follow-up outcome.
Trial-by-trial statistical model output not used for meta-analysis output. Treatment and control group counts used instead.	We specified in the protocol we would do this in the event that the process evaluation illustrated that the pooling approach to analysis would be unsound. However, we will not be doing this because running separate models prior to meta-analysis has not been factored into our power calculations and it is not necessary to do this to run a meta-analysis. We wish to present the results about between-trial variation in outcomes visually in any event.

	We will use basic descriptives from each site to populate the forest plot (treatment and control violence against the person offence counts, sample size).
Subgroup results reported as descriptive statistics for violence against the person prevalence and frequency.	We originally set out that we would conduct several subgroup analyses statistically. Instead we will report descriptive statistics for prevalence and frequency of violence against the person offences. This is to avoid issues with multiple hypothesis testing & low statistical power that would arise if we were to run subgroup analyses as models.
Not undertaking analysis for co-offender allocation to treatment.	The data we have will not allow us to link co-offenders so it will not be possible to assess the impact (or otherwise) of treatment allocation on co-offenders not in the trial or in the trial itself (as we do not have information on co-accused or co-defendants).
Not undertaking subgroup analysis for low, medium and high-risk groups	This measure was derived differently for each trial so unless the definitions are the same then the results from this would be uninformative . We will only present trial x risk-level descriptive statistics for treatment and control.
Not undertaking analysis of co-offending for violence against the person	We are interested in co-offending because of the focus on group involvement in crime – that is any group involvement in crime – and violence is too narrow a definition. When coupled with the low statistical power this analysis would have if it were focused on violence against the person alone, we did not think it was worth pursuing further.
Compliance analysis only at the individual level, not trial level	Our definition of compliance is whether those allocated to the treatment group actually received intervention of any sort and whether those in control remained in control and did not receive the intervention. While we can calculate a trial-level compliance measure from the ground up using individual compliance data, anything beyond initial contact is firmly into 'fidelity', that is, the 'how' of the intervention rather than 'whether the intervention was received or not'.
Changes in sample size calculations and assumptions for power simulation	Since the publication of the protocol we have been able to examine data from a randomly selected subsample of the

	cohort that has provided a sample of outcomes upon which to update our power simulations. This distribution of the primary outcome was used to inform the power simulations using a method previously described in the study protocol and early implementation reports. The method used in the SAP varied slightly to that employed previously in that the time 2 outcomes data was simulated in a different way for the control group. In the protocol and early implementation report, outcomes for time 2 (endline) were set to be identical to those of time 1 (baseline) for both the treatment and control groups, with time 2 for the treatment group then being adjusted to reflect the simulated effect size. The change from the initial run of the power calculations in the protocol to the SAP is that in the protocol the exact values for t1 (baseline) were copied to t2 (outcome) for the control group. In this version, the values for the control group were allowed to vary (randomly) but the overall distribution and mean remained the same. This better reflects the variation at case-level that we observe in reality. The simulation for the treatment group remains the same between the protocol and the SAP. Our updated power calculations in this SAP show that a sample size of 2,864 would detect a relative reduction of 20% in 79% of trials.
Changes to the control variables used in the main intention-to- treat analysis	We will be using *trial* fixed effects (7 trials in total) instead of delivery site/team fixed effects (5 sites in total) in the model. We will also be including month of randomisation, ethnicity and sex in the intention-to-treat model.

Table of contents

Introduction: a multicentre randomised controlled trial of focused deterrence

This document is the statistical analysis plan (SAP) for the Another Chance multicentre randomised controlled trial. It is an updated version of the analysis plan set out in the study protocol (Brennan et al., 2023), informed by what we now know about implementation and information that has influenced the conduct of the analysis for the project (Brennan et al., 2024).

Youth Endowment Fund 'Another Chance' Focused Deterrence programme

Focused deterrence interventions are based on the premise that combining (i) a credible threat of legitimate, swift, and targeted enforcement measures for those involved in violence (i.e., increasing the perceived risk of being caught), with (ii) the offer of comprehensive and tailor-made support, and (iii) a community-based message that violence will not be tolerated and that desistance is supported, delivered via individuals and organisations involved in delivery of the programme, will lead to reductions and/or desistance from further violent behaviour. There is a good deal of evidence for the effectiveness of these types of programmes when run and evaluated in a US setting and when evaluated using quasi-experimental designs with intervention at the area level (Braga et al., 2019). Since the publication of Braga et al.'s systematic review in 2019, randomised controlled trials of focused deterrence have been implemented (Olphin et al., 2024; Rosenfeld & Vaughn, 2024) but have been limited by small sample sizes and low statistical power (Olphin et al. 2024). One international exception is Operation Capesso, a large-scale RCT conducted in Victoria, Australia (Australian Institute of Criminology, 2023), but the results have yet to be published.

Most US implementations of focused deterrence have promoted the use of 'call ins' for gang members/group violence, which in turn has raised challenges for evaluation (Braga & Weisburd, 2014; Rosenfeld & Vaughn, 2024). Owing to differences in implementation and differences in the ability to compel attendance through the threat of revoking parole, UK tests of focused deterrence have not used call-ins extensively or successfully (Davies et al., 2016). In the YEF programme, as with most previous UK deliveries of interventions labelled as 'focused deterrence', the programme combines an offer of support, which requires voluntary engagement (i.e. cannot be compelled through parole conditions) and the threat of enforcement in the face of continued offending, which is conditional, but not voluntary. Furthermore, the US implementation has focused on adults while the lower age limit for the YEF programme is 14 years. In the UK, the treatment unit has tended to be individuals rather than groups, meaning that the intervention mechanism is likely to differ. This means there is a significant gap between the reported evidence base (Braga et al., 2019) and implementation in the UK. Again, the lack of rigorous evidence from the UK underpins the need for further testing.

Design summary

The Another Chance evaluation is a multicentre, two-arm randomized controlled trial involving seven interventions delivered across five cities in England. Each intervention has been designed following the YEF FD framework (Youth Endowment Fund, 2022) and, consistent with the general ethos of focused deterrence initiatives (Braga et al., 2019), adapted for local context and team structure. All seven interventions target people aged 14 years and over who have been, or are at risk of being, involved in group violence in the community.

The evaluation of this intervention is being conducted via a combined randomised controlled trial (RCT) and implementation and process evaluation (IPE) under a realist framework (Pawson & Tilley, 1997). Both of these are detailed in the trial protocol (Brennan et al., 2023). To summarise, the impact evaluation is a multicentre, two-arm randomised controlled trial implemented across seven trials in five sites. The implementation and process evaluation cover both what is being delivered, but also the views of those running the interventions and subject to them, as well as assessing compliance and fidelity in the programme. The integrated RCT/IPE means that measures from the IPE work will directly inform the analysis of the RCT but for the purpose of statistical analysis the main contributions from the IPE are (i) measures of compliance with allocation and (ii) fidelity to programme (for details of the realist approach see the YEF study protocol - the SAP is concerned with the statistical aspects of the study).

Each intervention involves an initial search for eligible population based on defined eligibility criteria, which consists of police data on offending, but may also involve input from statutory and/or voluntary sector partners. Typically, individuals are then considered by a panel who draw together intelligence about the individual to inform a strategy for intervention. Prospective cohort members are contacted individually by the delivery team, usually and preferably as a face-to-face meeting. At this meeting, a team member communicates to the individual that their offending behaviour has come to the attention of the police, leading to the meeting, and that continued offending will result in them being arrested and suffering legal consequences. They are also informed that, should they wish to avoid these consequences, they must desist from violence and the programme can offer them support to desist. If they choose to accept this offer of support, a follow-up meeting is arranged to identify and coordinate future support. If an individual chooses not to accept the offer of support, they may be referred to police for enforcement or disruption activity if they continue to be involved in violence. While eligibility criteria have been consistent since 2023, we are conscious of the risk of local ‘mission creep’ and net-widening. Our team is completing qualitative work to understand more about these issues. The protocol for this study was

registered as ISRCTN11650008 on 4th June 2023 and was updated in June 2025: https://doi.org/10.1186/ISRCTN11650008.

Target population

The population involved in this study is those aged 14+ years old with a history of offending and/or at high risk of offending (see protocol for detailed eligibility criteria, which differed slightly between trials). A further component of the YEF remit is to evaluate the most promising interventions to prevent violence, of which focused deterrence is one (Abt, 2019). However, the age profile for focused deterrence interventions typically is those in adulthood and arguably the intervention will have differing effects at different ages. To ensure that both components of the YEF remit can be addressed, a pragmatic decision was made to widen the age limit to include children and adults and to examine differences in intervention effects and experiences in children and adults as part of the formative evaluation (IPE). Below we set out the different sites, trial ID numbers and names by which the interventions are known locally.

Table 1: Local delivery sites & trial numbers

Trial number	Team	Site	Intervention name	Number randomised to date*
1	Leicester, Leicestershire and Rutland Violence Reduction Network	Leicester City	The Phoenix Programme	553
2	Greater Manchester Combined Authority	Manchester City	Another Chance Manchester	630
3	Nottingham Violence Reduction Partnership	Nottingham City & Nottingham County	Another Way	365
4	West Midlands Violence Reduction Partnership	Coventry City	CIRV Coventry high risk pathway (Trial 1)	214
5	West Midlands Violence Reduction Partnership	Coventry City	CIRV Coventry referral pathway (Trial 2)	343
6	West Midlands Violence Reduction Partnership	Wolverhampton City	CIRV Wolverhampton high risk (Trial 1)	292
7	West Midlands Violence Reduction Partnership	Wolverhampton City	CIRV Wolverhampton referral pathway (Trial 2)	212

Table note: * date as of 31/12/2024, meaning numbers reported here and those below will not match.

Design overview

Trial design (including number of arms)		**Multicentre, two-arm randomised controlled trial**
Unit of randomisation		Individual, stratified by offending frequency and age
Stratification variables **(if applicable)**		Number of offences in past two years (tertiles) Under/over 18 years: binary variable
Primary outcome	Variable	Perpetration of violent crime attributed to an individual in the 12 months following randomisation
	measure (instrument, scale, source)	N charges/cautions/convictions for violence against the person offences in local police records and PNC within 12 months of randomisation
Baseline for primary outcome	Variable	Perpetration of any violent crime in the 24 months prior to randomisation
	measure (instrument, scale, source)	N charges/cautions/convictions for any violence against the person offence in local police records and PNC within 24 months prior to randomisation
Secondary outcome(s)	Variable(s)	Involvement in co-offending in the 12 months following randomisation
	measure(s) (instrument, scale, source)	N charges/cautions/convictions for any offence in local police records and PNC where two or more perpetrators were charged/cautioned/convicted for a crime incident within 12 months of randomisation
Baseline for secondary outcome	Variable	Involvement in co-offending in the 24 months prior to randomisation
	measure (instrument, scale, source)	N charges/cautions/convictions for any offence in local police records and PNC where two or more perpetrators were charged/cautioned/convicted for a crime incident in 24 months prior to randomisation
Secondary outcome(s)	Variable	Time-to-offence up to 12 months following randomisation
	measure (instrument, scale, source)	Time in days between randomisation and first violent offence that resulted in a charge/caution/conviction in local police records and PNC
Baseline for secondary outcome	Variable	Perpetration of violent crime in 24 months prior to randomisation
	measure (instrument, scale, source)	N charges/cautions/convictions for any violence against the person offence in local police records and PNC within 24 months prior to randomisation

Sample size calculations overview

		Protocol	Randomisation
Minimum Detectable Effect Size (MDES)		20% relative reduction in violence against the person offences. (d = 0.08)[1]	20% relative reduction in violence against the person offences. (d=0.08)
Pre-test/ post-test correlations	level 1 (participant)	n/a[a]	n/a[a]
	level 2 (cluster)	n/a	n/a
Intracluster correlations (ICCs)	level 1 (participant)	n/a	n/a
	level 3 (cluster)	n/a	n/a
Alpha		0.05	0.05
Power		0.8	0.79[b]
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size		n/a	n/a
Number of participants	intervention	1,250	1,432
	control	1,250	1,432
	total	2,500	2,864[b]

Table notes: [a] We based our power calculations on simulated data so we do not use pre/post correlation in the same way. [b] Projected randomisation throughput based on simulations in Appendix 1

[1] Cohen's d is presented here for consistency with other YEF studies. However, those studies typically employ outcomes with an approximately normal distribution, such as the Strengths and Difficulties Questionnaire (Youth Endowment Fund, 2022). As *d* relies on the existence of symmetrical standard deviations, its use when data have a negative binomial distribution and large number of zero counts is not advisable because it does not align with the assumptions of either *d* or the underlying data distribution.

Our sample size calculations were completed *a priori* and published in the protocol (Brennan et al., 2023). The calculations were completed using R v4.0.4 (R Core Team, 2021) and the randomizr package (Coppock, 2024). The conclusion of those calculations, heavily informed by the use of the YEF focused deterrence framework across all sites, was that in order to achieve sufficient statistical power for the study it would be necessary to pool data across sites. Our code was published with the protocol, has been refined to automatically run and visualise all simulations from a single file, and is included in Appendix 1. Since the publication of the protocol we have been able to examine data from a randomly selected subsample of the cohort (described below) that has provided a sample of outcomes upon which to update our power simulations.

Data on outcomes were requested from sites to help inform the power calculations documented below. Lists of all cases sent for randomisation until the end of June 2024 were collated for each trial and 20 cases were randomly sampled from each of the seven trials. This was so that (i) the overall burden on site analysts was minimised and they could return the data quickly (each request requires a lookup on PNC and review of files to calculate the offence outcomes); and (ii) we could be more confident about inferences from the sample collected rather than asking sites to send back data on a self-selected sample. Site analysts were provided with this list and a set of instructions for returning information on outcomes for each case between 1/4/2023 and 31/3/2024.

The distribution of the primary outcome was used to inform the power simulations using a method previously described in the study protocol and early implementation reports. The method used here varied slightly to that employed previously in that the time 2 outcomes data was simulated in a different way for the control group. In the protocol and early implementation report, outcomes for time 2 (endline) were set to be identical to those of time 1 (baseline) for both the treatment and control groups, with time 2 for the treatment group then being adjusted to reflect the simulated effect size:

$$t_2 = \begin{Bmatrix} t_1, if\ x\ =\ control \\ t_1\ \times\ d, if\ x\ =\ intervention \end{Bmatrix}$$

where *d* is a relative reduction effect size. In this iteration, to better reflect natural variation on the outcome in the absence of an effect for the control group, the same data generating process was used for t1 and t2, which will generate the same group-level means and overall distribution for the control group, but will include case-level variation via the use of random variation in the data-generating process at the individual level (see R code in Appendix 1; code snippet pasted in footnote below).[2] The change from the initial run of the power calculations

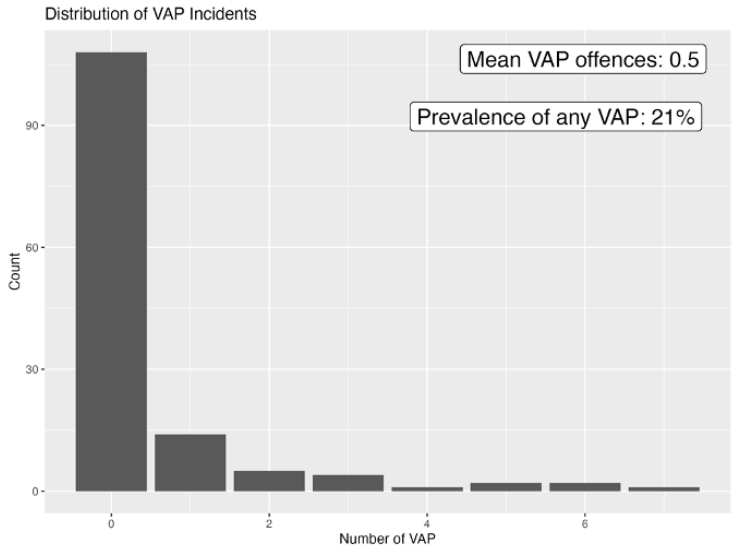
[2] # Create sample dataset
generate_sample_df <- function(n, d) {

in the protocol to the SAP is that in the protocol the exact values for t1 (baseline) were copied to t2 (outcome) for the control group. In this version, the values for the control group were allowed to vary (randomly) but the overall distribution and mean remained the same. This better reflects the variation at case-level that we observe in reality.

Primary outcome: Number of violence against the person offences

The sample outcomes data (n=140) are described in Figure 1. In 79% of cases, there was no violent offence in the follow-up period. The minimum number of violent offences was 0, the mean was 0.5 with a standard deviation of 1.28 and a maximum of 7 offences. The approximate distribution of the data was an over-dispersed Poisson distribution with an inflated rate of zeroes. Accordingly, the appropriate statistical model to estimate the effect of a randomised intervention is a zero-inflated negative binomial model.

Figure 1. Distribution of violence against the person (VAP) incidents



The resulting power achieved through different combinations of effect and sample size are visualised in Figure 2 below. As per the protocol, the viable effect size for the intervention is

```
  t1_outcome <- generate_t1_outcome(n)
  risk <- generate_risk(t1_outcome, 0.6)
  id <- seq_len(n)
  condition <- sample(c(0, 1), replace = TRUE, size = n)

  # Generate t2_outcome scaled by 'd' if condition == 1
  lambda_t2 <- ifelse(condition == 1, d * lambda, lambda)
  t2_outcome <- rpois(n, lambda = lambda_t2)

  tibble(id, condition, t1_outcome, risk, t2_outcome)
}
```

a relative reduction of 20% (d = 0.08). Under the parameters of the simulations, the analysis will require approximately 3,000 cases to achieve 80% statistical power. The likely achievable sample size under the terms of the projected number of cases will give 79% power to detect a 20% relative reduction in number of charge/cautions/convictions for violence against the person offences.

Secondary outcome: Co-offending

Outcomes data were also available for the secondary outcome, number of charge/cautions/convictions for a co-offence, and are presented in Figure 3. This outcome followed a similar Poisson-type distribution but with a higher frequency of zeroes (84% of cohort members had no co-offences in the follow-up period) and a lower mean number of offences: 0.2 with a standard deviation of 0.57 and a maximum of 4. The likely achievable statistical power under the terms of the projected number of cases is 48% power to detect a 20% relative reduction in the number of charge/cautions/convictions for co-offences in the cohort.

Secondary outcome: time-to-offence

We have not included a discussion of time-to-offence here as very few reoffended (21%) and the random sampling means that we had a mix of 'time at risk' for reoffending so there would be little consistency in exposure risk, with the net result meaning even fewer cases could be included in an event study at this point in time.

Figure 2. Statistical power across four effect sizes for primary outcome

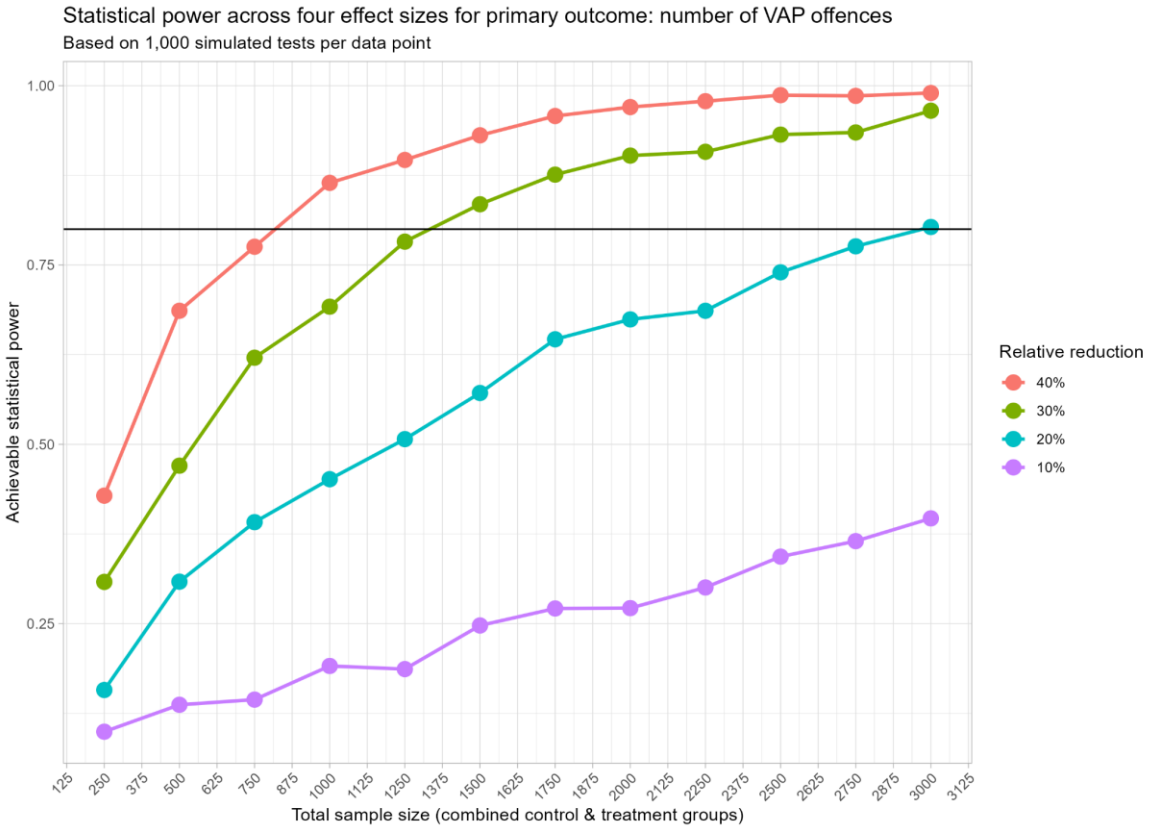
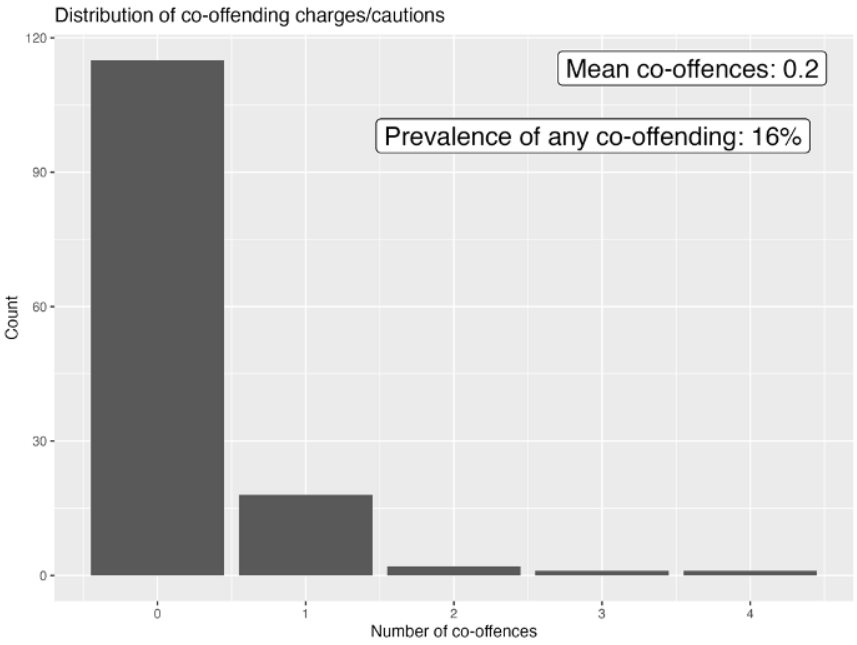


Figure 3. Distribution of co-offending charges/cautions/convictions



Analysis

Primary outcome analysis

Challenges with outcome measurement affecting this trial

The intention of this trial is to look at violent behaviour and whether the focused deterrence approach, as implemented, can reduce this. We know that official records of conviction undercount how much crime individuals commit (Cornish et al., 2025) but that the severity of violent crime is negatively correlated with reporting and, by extension, detection (Brennan, 2019). In order to ensure sufficient sample size, we relied on administrative data, routinely collected by police as our outcome measure.

The outcome defined in the protocol was that an offence was 'attributed to' an individual. This can be interpreted differently, supported by very different standards of evidence. At the most permissive interpretation, 'attributed to' could mean that, in police crime records or intelligence logs, an individual was named as a suspect for an offence but was not arrested or adjudicated for the offence. This low standard of evidence is susceptible to a range of biases, such as discrimination and, importantly, may not be a sufficiently specific (i.e. having fewer false positives) measure of the behaviour that the intervention is targeting. At the least permissive interpretation, 'attributed to' could mean that a person formally admitted to or was found guilty of an offence. The challenge facing the trial is to identify a suitable measure available in administrative records that is specific enough to truly reflect violent behaviour by an individuals and sensitive enough to ensure that violent behaviour is captured within the available 12 month follow-up period.

The first key challenge facing our team, and indeed anyone working on criminal justice evaluations in the UK at the moment, is that the justice system itself is not working efficiently. Specifically, there are significant backlogs in the court system that mean more than 70,000 Crown Court cases and 300,000 Magistrates' Court cases are yet to be heard (House of Lords, 2025; Criminal Court Statistics Quarterly, 2024). In 2024, the median times from offence to completion for a magistrates' court and a Crown court case were 182 days and 355 days, respectively (Commons Library, 2024). This means that criminal justice outcomes data are systematically 'right censored' (Sutherland, 2013) and we may not observe criminal justice outcomes for cases that need to go to court for trial or sentencing. Simultaneously, there has been an increased use of diversion from the justice system – codified in 'outcome 22' in the National Crime Recording Standards (Home Office, 2025) – but outcome 22 is not currently classed as a 'criminal justice outcome' but is instead recorded as 'no further action' and is not recorded consistently between forces (McFadzien & Phillips, 2021). This means that cases may not make it beyond diversionary approaches, even if guilt is admitted, meaning that those offences would not be 'counted' as cautions or convictions on an individual's criminal

record. To our knowledge, and barring a change in guidance or legislation, the issue regarding outcome 22 will remain.

The second issue relates to how we define our outcomes. If we focus solely on convictions (proven offending) then the systemic issues with the CJS set out in the previous paragraph seriously limit the overall number of offences. While that issue is equivalent across treatment and control groups, it has the overall effect of reducing power because in count models, the total number of outcomes observed is a key determinant of statistical power.

Finally, geographical coverage of data. Police National Computer (PNC) data is national - as the name suggests - which means it captures (re)offending in and out of a given police force area. If we rely on local police administrative data that is not PNC, we may not be able to see all out of area crimes, which can confound the analysis (more detail below).

Therefore, for the purposes of this trial, and in a clarification to the protocol, we will include violence against the person (VAP) offences that have been 'charged, cautioned or convicted', as our primary outcome. This is to try and ensure a higher overall number of in-scope outcomes for the trial, thus increasing statistical power. In terms of the timing of the data pull - this will be roughly six months after the final 12 month follow-up period has completed for those randomised to allow time for CJS outcomes to be realised and recorded in PNC. In the event we have to request data earlier, we realise this may induce some right-hand censoring but that will be equivalised across trial arms and the inclusion of charge data should increase the overall number of offences with outcomes (as discussed above). We will also request offences where a person has been arrested in addition to 'charged, warned or convicted' in order to allow us to look at whether there was a differential detection / sanction risk arising from being included in the FD scheme, or indeed a backfire for those in the control group.

We are applying for PNC access via the National Police Chiefs' Council (NPCC) in order to access arrests, charge, warning, caution and conviction data. In the event that we are unable to access PNC via this route then we will pursue this via the Ministry of Justice, with the drawback being that MoJ's version of PNC only includes crimes with caution or conviction outcomes (i.e. a smaller number of outcomes than the NPCC version).

Our data sources are preferred in the following order: 1. NPCC PNC data (includes charge and co-offending information; national in scale), 2. Police data (includes charge information; local intelligence on co-offending[3]) and 3. MoJ PNC data (does not include charge or co-offending information; national in scale). If access to the preferred source is not available, we will move to the next preferred source.

[3] See 'Secondary outcome analysis' section for more information on dataset considerations for co-offending.

The next two subsections (main statistical analyses using PNC and descriptive statistics using police data) assumes NPCC PNC will be available to us. Otherwise, we will use local police data for our main analysis, and present pooled descriptive statistics using MoJ PNC.

Main statistical analyses using Police National Computer data

As set out in the trial protocol (Brennan et al., 2023) the **main statistical analysis for this trial will take the form of a pooled approach**, where data from all trials will be combined in a single model, using fixed effects for trial-site specific heterogeneity and strata variables used in randomisation. The justification for that approach, as opposed to a trial *X* treatment interaction term, is that the same underlying intervention framework was planned for all trials. The use of trial fixed effects also reflects the likely difficulties with generalisability to other locations and individuals in those locations not in this trial (Senn, 2021; Youth Endowment Fund, 2021). The implementation of the YEF framework varied, which we have discussed in depth previously (Brennan et al., 2023, 2024), so **we will visually present between-trial variation via the use of meta-analytical techniques**. Forest plots will give a visual representation of the variation between trials, because each trial has its own treatment estimate, and how much each trial contributes to precision on the basis of sample size for each site will also vary (e.g. larger samples have larger symbols in the forest plot). We will use only basic descriptives from each site to populate the forest plot, rather than any formal statistical analyses of differences (see Appendix 4 for more detail).

The main results from the project will be based on the intention-to-treat (ITT) analysis for the primary outcome using pooled data. The statistical model for the intention-to-treat analysis is set out below in equation 1. The model incorporates variables to account for between-trial and over-time variation, as well as variables used for stratification and person-level control variables.

$$Y = \alpha + \beta 1[treatment] + \beta 2[risk] + \beta 3[adult] + \beta 4[trial] + \beta 4[time] + \beta 6[ethnicity] + \beta 7[sex] + \epsilon$$

(Eq.1)

In the equation, *Y* is the outcome – in this case, and for our primary analysis, a count variable measuring the number of violent offences that have been 'charged, cautioned or convicted' in the twelve calendar months following randomisation of that person. The analysis approach will be based on count outcomes – we intend to use a *zero-inflated negative binomial model*, based on the observation that many individuals in the test sample did not have further

offences in the follow-up period of 12 months post-randomisation and the outcome was over-dispersed. The model variables are:

- $\beta 1[treatment]$ will be a binary variable where 0=control and 1=treatment and the coefficient from this variable in the model will be the focal result for the project.
- $\beta 2$[*risk*] is one variable that was used for stratification in each trial. This will be included in the analysis as n-1 dummy categories, with the reference category being the category with the largest number of observations from low, medium or high offending frequency, which was trial-specific (note that not pre-specifying which category now will not affect the results).
- $\beta 3[adult]$ will be a binary variable for whether an individual is aged 18 years and older (=1) or a child (=0) at the time of randomisation.
- $\beta 4$[*trial*] will be trial fixed effects - dummy variables for n-1 trials, again with the trial with the largest number of observations overall as the reference category. We include this variable because we know, *a priori*, that trials will differ in their eligibility criteria and selection processes, so we need to parcel out this variation in our analysis.
- $\beta 5$[*time*] will be a variable that captures the year and month since the first randomisation. This measure will capture seasonal variation as well as any esoteric shocks during the delivery period. This will be entered as dummy variables for each year-month, but if there are model convergence problems then we would aggregate this to year-quarter.
- $\beta 6$[ethnicity] will be n-1 dummy variables with White as the reference category (since this is the largest group, to be consistent with how other categorical variables will be entered in the model), and dummies for Asian, Black, Mixed Heritage, and Other ethnic minority groups. It may be necessary to combine ethnic categories with small sample sizes. Based on existing data from the previous reports, 50% white, 30% black, 9% Asian, 9% Mixed, 1% Other (Brennan et al., 2024), we anticipate the need to combine Other with Asian and/or Mixed Heritage groups for analysis.[4]
- We will also include a variable for offender sex $\beta 7$[*sex*]. This would be a binary variable with 0=male and 1=female, determined by sex at birth if possible to determine this from available data. It is important to note that the highly imbalanced nature of this variable creates a risk for model convergence.
- ϵ is the error term for the model.

We are including both ethnicity and sex in the analysis model because both are associated with differential risks of offending. Including them has the statistical benefit of increasing

[4] This is so that we do not end up with very small groups in the analysis model - that leads to problems with model convergence and very large standard errors and confidence intervals.

power (by reducing between-group variation) even if these variables were not part of the randomisation strategy (Bartlett, 2021; Grizzle, 1982). However, in cases where the variable is highly imbalanced, such as sex, inclusion of the variable may present modelling challenges. Again, as previously specified in the protocol, we will use robust standard errors and calculate 95% confidence intervals based on those. The standard error adjustment will be via the lm_robust package from the 'estimatr' package (Blair et al., 2024). We know that standard error adjustment is sensible given that this helps in the event of model misspecification and in the face of heterogeneous treatment effects (Cunningham, n.d.; Snijders & Bosker, 2012; White, 1980). Our model specification will be the same for primary and secondary outcomes. All intention-to-treat and secondary analyses will be performed using the R statistical software. Analysis code is provided in Appendix 4.

Descriptive statistics for outcomes after 12 months - Local police records

We will be using the Police National Computer (PNC) data set for our main analysis. Our rationale for using local police data for outcomes after 12 months is that offences that occur within a police force area are first collected by local police forces and become available in police records more rapidly. Police records are then collated by the Ministry of Justice and form part of the Police National Computer (PNC) data set.

As a data source, the PNC has an advantage over local police records since the outcome unit of interest is individual offending, whereas local police records are based on offending within a place (i.e. likely to be geographically bounded). Therefore, offending in area B by an individual who is involved in a trial in area A may not be included in a data set of local police records (if area B is not a study area). Moreover, some of this cohort are likely to be involved in County Lines offending (which involves travel and, potentially, offending outside one's home area) and the intervention can seek to reduce this type of exploitation and help individuals disengage from involvement in a County Lines network. Any missing out-of-force offences will potentially bias the data and, if the intervention is successful, dilute the effects of the intervention (i.e. failure to capture out of area offences could undercount offending in the control group). A recent linkage of police records and PNC found that around 14% of offences by individuals up to age 21 years occurred outside their resident police force area (Boyd et al., 2022) and did not appear in the records of the former.

However, as there is an additional administrative stage to PNC collation and additional data-cleaning is undertaken, PNC data take at least six months to become available. Allowing this time lag will also allow outcomes (e.g. convictions) to accrue within the PNC.

The trial is being undertaken at a time of intense interest in the use of focused deterrence as an intervention to prevent violence. Accordingly, quick deployment of preliminary results is desirable to inform policy being developed in the coming months. To be able to respond to

this need we will use local police records from each police force area to collate descriptive statistics on offending outcomes.

Given the interest in obtaining results as soon as possible, the project funders have requested that preliminary tests of the study hypotheses be undertaken using local police data to However, producing indicative results by doing preliminary tests of the study hypotheses using local police data (ahead of the definitive results based on PNC data), has methodological implications of testing hypotheses more than once, most notably the increased risk of false positive results. A solution to this risk, known as 'alpha spending' is to 'share' the p-value threshold between analyses, increasing the severity of the main statistical test (Demets & Lan, 1994). This has implications for statistical power. For example, sharing the alpha equally across preliminary and main results would reduce the statistical power in our simulation (where n = 2,864 and relative risk = 0.8) from 0.786 to 0.705. Spending less on the preliminary analyses, such as a 25/75 split, reduces the statistical power to 0.761, equivalent to the contribution of 350 cohort members. A more common approach is to split the alpha in a more extreme way, e.g. 2/98 (where the alpha is distributed 0.001/0.049). In this case, the power of the main statistical test is barely affected (0.783), although the power of the preliminary analyses is considerably reduced: around 44% meaning a high type 2 error rate.

A less quantifiable risk presented by undertaking preliminary analyses in an environment where results are eagerly awaited is that preliminary analyses are incorrectly taken as definitive. This is particularly problematic as it allows two sets of results to exist from which users may choose their 'preferred' result even in the presence of statements about the primacy of one set of results. This has the potential to harm the scientific literature and the credibility of the trial.

In light of these risks, a compromise we have reached is to conduct only present pooled descriptive statistics of the treatment and control outcomes using local police data in advance of the main statistical analyses that will use Police National Computer data. As for local police data, we will present only pooled descriptive statistics of the treatment and control outcomes and not statistical tests. Qualitative findings from the IPE may be provided in advance of these final impact results.

In the event that NPCC PNC is not available to us, we will use local police data for our main analysis (see previous subsection describing our preferred data source). With the caveat that there may be some out of area data missing from local records as per Boyd et al. (2022).

Secondary outcome analysis

We will follow the same model specification as for the primary outcome using Equation 1. The equation and covariates will be the same in all cases for co-offending (for which we expect to apply a *zero-inflated negative binomial model*, similar to the primary outcome) and time-to-event models (for which we expect to use a Cox proportional hazards model, as outlined in the protocol). We will report metrics from the specific analysis such as a hazard ratio for the survival model. Analysis code is provided in Appendix 4.

For the co-offending outcome, it is possible that we will have to rely on local administrative police datasets. To our knowledge, these datasets will include local intelligence on co-offending, but with the caveat that out of area offences may be missing. However if this is available to us via NPCC PNC (our preference for analysing the primary outcome) we will pursue this route.

Subgroup analyses

We know based on our power calculations and as reported in the protocol that the trial is not powered for subgroup analyses. As stated on page 101 of our YEF evaluation protocol (Brennan et al., 2023):

> "The study will not be powered for subgroup analyses that rely on null hypothesis statistical testing because the approach to conducting subgroup analyses in trials is to run models as interactions between group and treatment. This typically means requiring a sample size ~x4 times larger than that for the main effect analysis…[W]e will still undertake subgroup analyses for the different risk groups in each site [sic; ''trial' is the correct term], and will report point estimates and confidence intervals but will not report or share p-values with YEF or sites. This will allow an assessment of the direction and magnitude of effect without the bias of 'statistical significance' that we know, *a priori*, will be very likely due to chance and not a real effect, even if pre-specified."

At the time of writing and with more than 2,800 participants randomised, it is still the case that subgroup analyses would be underpowered, largely but not exclusively owing to low base rates of VAP outcomes. As such, undertaking numerous statistical tests for different subgroups risks chance differences arising through subsetting data or multiple testing and confusion about which results are reliable. To avoid this, we will undertake exploratory subgroup analysis (via the use of meta-analytical techniques; see Appendix 4) for different subgroups based on the classification by individual or delivery characteristics. For all

subgroup analyses we will report descriptive statistics on the prevalence and frequency of VAP offences in the 12m post-randomisation. This will consist of the mean / median depending whether prevalence or frequency (respectively), the range of offences for the frequency and a measure of dispersion such as standard deviation. We will present these visually in a forest plot, and will not be undertaking statistical analyses of differences.

Risk subgroups

We pre-specified that we will look at differences for different risk groups in each trial, without disclosing p-values. So for example we will report on the treatment/control differences for 'high risk', 'medium risk' and 'low risk' groups through sub-setting the data and reporting the descriptive statistics as set out above. The categories of high, medium and low risk were set based on tertiles of offending frequency from approximately the first 10% of the cohort in each trial. As a result, high, medium and low are not comparable across sites. Therefore, we will produce descriptive statistics across trial and risk level, but will not undertake statistical analyses of differences.

Adult versus child subgroups

Based on previous research and following discussions with sites, we would expect to see differential impacts for young people (under 18 at the point of randomisation) versus adults (aged 18+ at the point of randomisation) and across different levels of future risk of reoffending (Kilkelly, 2023). As above we will report treatment/control differences across adult/child subgroups and risk level, but will not undertake statistical analyses of differences.

Ethnic subgroup analysis

The YEF guidance on subgroup analysis makes specific mention of 'race [sic] and ethnicity'.[5] We have routinely collected data on ethnic groups for participants but as above we know the trial is not powered for this analysis. We will report differences in treatment versus control means for each of the five ethnic groups classified in our data using the unadjusted prevalence of VAP post-randomisation and VAP counts.

Subgroup analysis by delivery approach

As reported in the early implementation findings report (Brennan et al., (2024); and see Appendix 3 on Context-Mechanism-Outcome configurations) the delivery approach varies between trials. Based on the team's fieldwork, there is sufficient contextual difference between the police-delivered CIRV trial and other trials that a police-led vs not police-led

[5] A guide to race and ethnicity terminology and language | The Law Society

comparison is justified. Specifically, having police officers act as enforcers *and* those offering support is a marked difference to other approaches where these roles are differentiated (or enforcement is largely absent). We will conduct a police-led vs not police-led comparison to assess whether there are statistically distinguishable differences between CIRV (Wolverhampton and Coventry) and other delivery models. To create the police-led vs not police-led cases for analysis we will use trial IDs to create new grouping variables where police-led=1 and not police-led=0. We will report treatment/control differences for each group using the unadjusted prevalence of VAP post-randomisation and VAP counts.

Trial-by-trial analysis

In our protocol (Brennan et al., 2023, p. 17) we set out that we would use the statistical output from a trial-by-trial analysis to populate a forest plot in the event that there is 'insufficient homogeneity' in delivery to combine all sites in a single ITT model, acknowledging that this would be 'likely to suffer from insufficient statistical power'. This approach was also premised on the idea that there were differences between sites in terms of eligibility criteria and strata variables used for randomisation. We also explained in the early implementation report (Brennan et al., 2024, p.114) that the evaluation model for Trial 2 in Coventry and Wolverhampton (Trials 5 and 7 in Table 1 above) is different from the other five trials and is a risk to the legitimacy of pooling data from the seven trials into a single multi-centred trial.

In a change from the protocol (detailed in the change log at the beginning of this document), we will present a forest plot to illustrate between-site heterogeneity on the basis of (i) the unadjusted prevalence and (ii) VAP counts for the treatment and control groups in each trial. This will allow us to show the extent of variation between sites that the ITT analysis averages out.

Furthermore, we will not conduct trial-by-trial statistical analyses that adjust for additional strata used in randomisation. This is because in the case of Nottingham, where urban/rural cases were supposed to be continuously randomised as separate strata, this in fact only happened in the three randomisation batches. This meant that only 15 / 365 cases to date (~5%) were randomised in this way. Second, for Leicester, the initial batch were randomised according to whether participants were under statutory supervision or not as part of an agreement for local probation/youth justice services to provide an in-kind contribution to the project. This was to ensure those working for statutory services had sufficiently high caseloads. Following the initial randomisation, the throughput of cases meant that an anticipated issue with low caseloads did not materialise. As such, it was not necessary to carry on the stratification.

Further analyses

Our further analysis intends to understand the heterogeneity of treatment effects in this sample. This will be conducted on an exploratory basis using machine learning techniques such as Classification And Regression Trees (CART) (Krzywinski & Altman, 2017) to identify for whom the approach may be most effective for. The results from this analysis will form a separate publication from the end point YEF report. We are only beginning to work through the parameters for this analysis now so cannot report more detail here.

In the protocol (p.98) we stated:

> "Information on co-offending will be captured as a secondary outcome. It may be possible to examine patterns in the treatment allocation of co-offenders within the outcomes data set. This will allow us to describe and potentially adjust for the relationship between treatment condition and co-offender treatment condition (or absence from the cohort entirely). The ability to do this rigorously will form part of the pilot activity and will inform the statistical analysis plan."

However, on the basis of the data we have (and will have) available, it will not be possible to tie together the allocation groups and who co-offenders are so will not pursue this further.

Interim analyses and stopping rules

As set out in our protocol we conducted interim analyses six months after the trial began as part of routine reporting to YEF and to fulfil our stopping rule check. The interim analysis looked at the direction and magnitude of impact without the use of statistical analysis because of the problem of 'alpha spending' (Meurer & Tolles, 2021). That is, we were focused on absolute differences between treatment and control conditions and had two rules in place to decide on whether to stop the intervention or stop the trial and roll-out the intervention:

- "In the event that the average impact is negative [equal to or greater than 10 percentage points] then the study will pause intake for one month to allow for options regarding progression to be tabled and agreed upon.

- If the average impact is positive, the threshold for roll-out to all participants will be higher. Reoffending prevalence would have to be 15 percentage points lower in the treatment group than control participants (e.g. 30% in control, 15% in treatment). This

asymmetry reflects that we want to be more cautious (more sensitive) to negative effects than positive ones" (Brennan et al., 2023, p. 102).

Our early implementation analyses (Brennan et al., 2024) - reported in Appendix 5 - indicate that potential harm and potential benefit were within the permitted threshold. For violence against the person, the absolute difference in the prevalence of reoffending between groups was *four* percentage points (21% vs 17%). Therefore, there was no need to stop the ongoing full implementation of the trial on the grounds of harm or benefit. This position is supported by the absence of any qualitative evidence that the intervention or the evaluation is causing harm.

Longitudinal follow-up analyses

Longitudinal analyses of the trial cohort is dependent on follow-on funding, but as previously set out in the protocol we will prepare for 24 and 36 month post-randomisation follow-ups. In this event, the analysis models will mirror those used for the main study based on Eq.1. The long-term follow-ups would be based on the intention-to-treat analysis set out for the primary and secondary outcomes set out above, with adjustment by outcome for multiple testing using the (Benjamini & Hochberg, 1995) false discovery rate correction. We will also include compliance analysis in the 24- and 36-month post-randomisation follow-up analyses, following the approach set out here.

Imbalance at baseline

As per our protocol and as reported in the early implementation report, we assessed balance *to date* - as randomisation is still happening - using an assessment of the means and distributions of variables in both groups for each site but have not statistically tested for differences between groups. Given appropriate randomisation procedures were followed, any difference between groups at baseline will be due to chance (Brennan et al., (2023, 2024); see also CONSORT guidance).[6] In Table 3 we report on the balance on participant characteristics at baseline for the 1,676 randomised during the early implementation phase. For the final report, we will report a table of means and standard deviations for all stratification variables, trials/sites, participant sex, ethnicity and prior offending, split by treatment and control groups, the data for which will be available at the end of the trial. For the prior offending variable, following our protocol (p.98), we will show the distribution of

[6] http://www.consort-statement.org/checklists/view/32-consort/510-baseline-data.

this variable by treatment and control group as a graph. Similar to the early implementation phase and as justified above, we will not be conducting any statistical tests to assess balance. In the event of imbalance and prior to undertaking any outcome analyses, we will include imbalanced variables. Imbalance will be defined as a standardised mean difference greater than our target effect size of a 17% treatment/control difference. That is, we have planned the trial to detect a 17% difference between treatment and control - if we see baseline imbalances that large or larger then we will include those variables in the analysis as controls, provided that they were measured prior to treatment allocation and are otherwise unaffected by treatment allocation.

Intervention and control group balance

Twenty-seven batches of cohort members were randomised between May and December 2023, totalling 1,676 individuals. Table 3 shows that allocations to treatment and control in each trial have been successful, in that there have been no randomisation failures in terms of equal allocation / by level of risk stratification, i.e. we planned for 1:1 allocation to treatment and control, and that is what we observe.

Table 3. Distribution of random assignment and stratification variables

	Control	**Treatment**	**Overall**
N	**847**	**829**	**1676**
Site (%)			
Coventry	163 (19.2)	156 (18.8)	319 (19.0)
Leicester	223 (26.3)	220 (26.5)	443 (26.4)
Manchester	195 (23.0)	199 (24.0)	394 (23.5)
Nottingham	114 (13.5)	115 (13.9)	229 (13.7)
Wolverhampton	152 (17.9)	139 (16.8)	291 (17.4)
Trial (%)			
Coventry Trial 1	92 (10.9)	92 (11.1)	184 (11.0)
Coventry Trial 2	71 (8.4)	64 (7.7)	135 (8.1)
Leicester trial 1	223 (26.3)	220 (26.5)	443 (26.4)
Manchester trial 1	195 (23.0)	199 (24.0)	394 (23.5)
Nottingham trial 1	114 (13.5)	115 (13.9)	229 (13.7)
Wolverhampton trial 1	102 (12.0)	102 (12.3)	204 (12.2)
Wolverhampton trial 2	50 (5.9)	37 (4.5)	87 (5.2)
Adult = yes (%)	457 (70.1)	433 (68.7)	890 (69.4)
Offence frequency (%)			
Low	209 (32.1)	188 (29.8)	397 (31.0)
Medium	223 (34.2)	213 (33.8)	436 (34.0)

High*	220 (33.7)	229 (36.3)	449 (35.0)

Table note: *Cut points for low, medium and high were trial specific.

Missing data

Since our protocol was published we have begun the process of requesting outcome data from the Ministry of Justice. This is a central government department responsible for the administration of courts, prisons, probation services and attendance centres. The Ministry of Justice analytical services directorate has access to the Police National Computer (PNC) which is administered by police forces across England and Wales and has been used extensively for reconviction studies (Sutherland, 2013). Our decision to pursue central data access was in part driven by the desire to have a single point of contact and source for the outcome data to ensure consistency and reduce complexity with data requests. Given this shift to a centralised administrative database, we anticipate low levels of missingness on outcomes from PNC data. That said, we know *a priori* that PNC data will not include out-of-court disposals or community resolutions because neither of these lead to criminal records. This means that lower level offences may be resolved with these disposals and those outcomes not included on individuals' PNC records. We would therefore need to exclude these as unproven offences but as this will be equivalent between treatment and control conditions the impact should be symmetrical across arms (i.e. unbiased).

In the event that we do encounter missing data on outcomes, the strategy used depends on the extent of missingness and whether we know or believe missingness to be Not Data Dependent (NDD), Seen Data Dependent (SDD) or Unseen Data Dependent (UDD) (Hand, 2020). NDD is ignorable because it is entirely unpredictable (i.e. random) and does not depend on the values of the data, meaning that complete case analysis will suffice. For SDD this means we know or understand that missingness is determined by a covariate or covariates - if these are included in analysis then missingness is 'as random conditional on those variables'. For Unseen Data Dependent (UDD) the missingness depends on the value of the missing variable or on values of data we do not have. In the case of this trial, it might be that everyone with five offences or more was missing. It is not possible to test whether data are UDD - we have to make judgments based on our knowledge of the data and the context of the work.[7]

To reiterate, we believe the risk of missingness for outcomes is greatly diminished because we are using a national policing database. However, in the event there *is* missing data on outcomes, and again as previously set out in our protocol, we will undertake the following

[7] These definitions are akin to the more widely used but less accessible MCAR, MAR and MNAR terms used for missing data.

steps for the primary outcome depending on the nature and extent of the missingness. If there is missingness greater than 5% of the total sample we will explore missingness across trial arms. To assess whether there are systematic differences between those who we do not observe outcomes for – and whether factors should be included in analysis – we would model missingness at follow-up as a zero/one outcome as a function of baseline covariates, including treatment (i.e., following Equation 1). This will tell us whether a specific variable or variables predicts missingness.

1. For less than 5% missingness overall a complete-case analysis should suffice, regardless of the missingness mechanism (Education Endowment Foundation, 2018).

2. For between 5-10% missingness our default would be to check results using approaches that account for missingness that rely on the weaker SDD assumption, building the SDD conditioning variables from our initial work predicting missingness. If there was systematic missingness of predictor variables, for example, we use full information maximum likelihood (FIML) (Education Endowment Foundation, 2018); for a discussion of FIML vs MI see Allison, 2012).

3. In the event that baseline data are unavailable or missing for individuals, those individuals would be included in the outcome analysis via FIML, rather than sacrifice statistical power through excluding them.

4. In the event of more than 10% missing data on outcomes, assuming that is equally distributed across treatment and control arms, we would first use min-max imputation. That is, we impute the lowest and highest values as missing values to give the extreme range of impact that missingness could have. This obviously assumes that all missing cases fit into these extremes, which is unlikely, but it also provides an empirical floor and ceiling for treatment effect differences in light of missingness. We would then also report the FIML result from step two as the main result but use the min-max approach as sensitivity analyses.

5. For between 11-20% missing data we would explore the feasibility of using multiple imputation referring to guidance on missing data analysis (Allison, 2002; Rioux & Little, 2021). Similarly, in the event that we believe data are UDD then imputation would be necessary to model the result. If imputation is pursued, we will follow guidance such as (Austin et al., 2021).

6. For more than 20% missing data on outcomes, particularly if asymmetrical across treatment arms, the risk of bias is substantial (Sterne et al., 2019), and it would not be clear which way the bias would run (e.g. towards or away from zero). We would use

multiple imputation with fifty imputed datasets to estimate the study results but will caveat substantially regarding interpretation.

Compliance

Compliance means 'compliance with allocation' (Ariel et al., 2022) - that is, did those allocated to treatment actually receive intervention of any sort and did those in control remain in control and *not* receive the trial intervention? For this trial, *compliance will be measured at the participant level*. The compliance measures created will be used to estimate the complier-average-causal-effect (CACE; Angrist and Pischke (2009)).

Participant level compliance will be measured by assessing each participant against data about whether or not they received the intervention message based on process data collected from site teams. That is, were participants made aware of the programme via an offer of support and/or deterrence messaging? This will be coded as a 1 where we have evidence they did, and 0 if they did not. This will mean for some individuals who 'timed out' in terms of eligibility or were uncontactable/untraceable but were allocated to treatment, they will receive a zero.

For those in the control group, we will use the same coding approach in relation to offers of support or deterrence activity by the site teams only (i.e. ignoring if other teams intervened or other services were working with the individual). If control participants received an offer of support and/or deterrence messaging we will code that as "1". That will then mean in the allocation compliance measure they will be coded as "0" i.e. *non*-compliant with control. (The question might arise about other support sought out or police deterrence/ enforcement activity for control group members being evidence of 'non-compliance'. However, both treatment and control group members can seek additional support at any time and both can be the focus of additional police activity, so this is equivalised across groups.) The cross-tabulation between the compliance variable in treatment and control groups that will give us our overall level of compliance measure as a percentage at the trial level.

For those missing all data on compliance with support or deterrence, we will assume compliance to allocation and that will inform the CACE analysis (below) as the definitive result. We will assess the impact of this assumption by also using the opposing assumption of non-compliance in the CACE. This min-max approach will give us the extreme range that the CACE analysis can have in the presence of missing compliance data.

The compliance measures created will be used to estimate the complier-average-causal-effect (CACE) estimate using 2SLS (Angrist & Pischke, 2009). The treatment indicator will be the instrument (1=treatment, 0=control). Following the example set out in (Heiss, 2020) the treatment indicator is *relevant* as being told about the offer/threat is highly likely to lead to

changes in behaviour. It is *exclusive* because the only way that the programme can cause changes in violence is through first making the participant aware of the programme and that only happens via being made aware via random allocation. And finally it is *exogenous* because being told about the programme is not related to other things that might cause changes in offender behaviour or anything else in the model, because it is random.

To estimate the CACE we will use the iv_robust command from the estimatr package (Blair et al., 2024). We will estimate the result using robust standard errors. Analysis code is provided in Appendix 4.

Presentation of outcomes

We will report on the offending outcomes in this trial on the basis of (a) raw offence counts per treatment group (in a table and histogram similar to Figures 1 and 3); (b) relative risk (RR) from the zero-inflated negative binomial regression models specified above. For time-to-offence, we will report (a) average (mean) time to first violent offence per treatment group (in a table and graph illustrating distribution); (b) hazard ratio (HR) from the Cox proportional hazards model. For RR and HR, these will be reported with the standard errors and test statistics to three decimal places (rather than thresholds for p-values) and frequentist 95% confidence intervals for the relative risk based on multiplying the point estimate by 1.96 times the standard error.

Bibliography

Abt, T. (2019). *Bleeding out: The devastating consequences of urban violence—And a bold new plan for peace in the streets*. Hachette.

Allison, P. D. (2002). *Missing Data*. SAGE Publications. https://doi.org/10.4135/9781412985079

Allison, P. D. (2012). *312-2012: Handling Missing Data by Maximum Likelihood*. SAS Global Forum 2012. https://statisticalhorizons.com/wp-content/uploads/2022/01/MissingDataByML.pdf

Altman, D. G., & Bland, J. M. (2011). How to obtain the P value from a confidence interval. *BMJ*, *343*, d2304. https://doi.org/10.1136/bmj.d2304

Angrist, J. D., & Pischke, S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ariel, B., Bland, M., & Sutherland, A. (2022). *Experimental Designs*. SAGE.

Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing Data in Clinical Research: A Tutorial on Multiple Imputation. *Canadian Journal of Cardiology*, *37*(9), 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

Australian Institute of Criminology. (2023, October 16). *Victoria Police-led Operation Capesso is recognised for its efforts in reducing crime among Melbourne's most serious violent offenders*. https://www.aic.gov.au/media-centre/news/victoria-police-led-operation-capesso

Barnes, J. C., TenEyck, M. F., Pratt, T. C., & Cullen, F. T. (2020). How Powerful is the Evidence in Criminology? On Whether We Should Fear a Coming Crisis of Confidence. *Justice Quarterly*, *37*(3), 383–409. https://doi.org/10.1080/07418825.2018.1495252

Bartlett, J. (2021, August 6). *Stratified randomisation*. The Stats Geek. https://thestatsgeek.com/tag/stratified-randomisation/

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blair, G., Cooper, J., Coppock, A., Humphreys, M., & Sonnet, L. (2024). *estimatr: Fast Estimators for Design-Based Inference* (v1.0.2). https://declaredesign.org/r/estimatr/

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2022). *Comprehensive Meta-Analysis v4* [Computer software]. Biostat. https://meta-analysis.com/

Boyd, A., Teyhan, A., Cornish, R., Croft, J., Thomas, R., Brennan, I. R., & McLeod, J. (2022). The potential for linking cohort participants to official criminal records: A pilot study using the Avon Longitudinal Study of Parents and Children (ALSPAC) [version 2; peer review: 2 approved]. *Wellcome Open Research*, *5*(271), 1–23. https://doi.org/10.12688/wellcomeopenres.16328.2

Brader, C. (2025). *Reducing the crown court backlog*. House of Lords Library. https://lordslibrary.parliament.uk/reducing-the-crown-court-backlog/

Braga, A. A., MacDonald, J. M., Douglas, S., Wade, B., & Struhl, B. (2024). Focused deterrence, strategic management, and effective gun violence prevention. *Criminology & Public Policy*, *23*(4), 919–946. https://doi.org/10.1111/1745-9133.12682

Braga, A. A., & Weisburd, D. L. (2014). Must we settle for less rigorous evaluations in large area-based crime prevention programs? Lessons from a Campbell review of focused deterrence. *Journal of Experimental Criminology*, *10*(4), 573–597. https://doi.org/10.1007/s11292-014-9205-8

Braga, A. A., Weisburd, D., & Turchan, B. (2019). Focused deterrence strategies effects on crime: A systematic review. *Campbell Systematic Reviews*, *15*(3), e1051. https://doi.org/10.1002/cl2.1051

Brennan, I., Simanovic, T., McFarlane, P., Sutherland, A., Graham, W., Powrie, M., Larigkou, I., & Holmes, E. (2024). *Focused Deterrence Randomised Controlled Trial: Early Implementation Report*. Youth Endowment Fund.

Brennan, I., Simanovic, T., Sutherland, A., McFarlane, P., & Graham, W. (2023). *ACF1-Focused-deterrence-Evaluation-protocol*. Youth Endowment Fund. https://youthendowmentfund.org.uk/wp-content/uploads/2023/07/ACF1-Focused-deterrence-Evaluation-protocol.pdf

Brennan, I. (2019). Victims of serious violence in England and Wales, 2011-2017. College of Policing. https://library.college.police.uk/docs/college-of-policing/Victim_of_serious_violence.pdf

Case management guidance—How to use out-of-court disposals—Guidance—GOV.UK. (n.d.). Retrieved 18 June 2025, from https://www.gov.uk/guidance/case-management-guidance/how-to-use-out-of-court-disposals

Commons Library (2024). Court statistics for England and Wales. Commons Library. https://researchbriefings.files.parliament.uk/documents/CBP-8372/CBP-8372.pdf

Coppock, A. (2024). *randomizr: Easy-to-Use Tools for Common Forms of Random Assignment and Sampling.* (v1.0.0). https://declaredesign.org/r/randomizr/

Cornish, R., Teyhan, A., Tilling, K., Macleod, J. and Brennan, I. (2025) "Measuring serious violence perpetration: comparison of police-recorded and self-reported data in a UK cohort", International Journal of Population Data Science, 10(1). doi: 10.23889/ijpds.v10i1.2391.

Criminal court statistics quarterly: October to December 2024. (n.d.). GOV.UK. Retrieved 18 June 2025, from https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-october-to-december-2024

Cunningham, S. (n.d.). *Causal Inference The Mixtape*. Retrieved 29 November 2024, from https://mixtape.scunning.com/

Davies, T., Grossmith, L., & Dawson, P. (2016). *Group Violence Intervention London: An Evaluation of the Shield Pilot*. MOPAC Evidence & Insight, Greater London Authority. https://www.london.gov.uk/sites/default/files/gvi_london_evaluation270117.pdf

Demets, D. L., & Lan, K. K. G. (1994). Interim analysis: The alpha spending function approach. *Statistics in Medicine*, *13*(13–14), 1341–1352. https://doi.org/10.1002/sim.4780131308

Education Endowment Foundation. (2018). *EEF_statistical_analysis_guidance_2018*. Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF_statistical_analysis_guidance_2018.pdf

Fazel, S., Hurton, C., Burghart, M., DeLisi, M., & Yu, R. (2024). An updated evidence synthesis on the Risk-Need-Responsivity (RNR) model: Umbrella review and commentary. *Journal of Criminal Justice*, *92*, 102197. https://doi.org/10.1016/j.jcrimjus.2024.102197

Grant Obo, K., & Heydari, A. (2025). *Outcome 22 (O22) Guidance 2022* [Guidance]. National Police Chiefs Council. https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/publications-log/criminal-justice/2023/npcc-outcome-22-guidance-2022.pdf

Grizzle, J. E. (1982). A note on stratifying versus complete random assignment in clinical trials. *Controlled Clinical Trials*, *3*(4), 365–368. https://doi.org/10.1016/0197-2456(82)90026-5

Hand, D. J. (2020). *Dark Data: Why What You Don't Know Matters*. Princeton University Press. https://doi.org/10.1515/9780691198859

Heiss, A. (2020, October 28). *Complier average treatment effects*. Program Evaluation. https://evalf20.classes.andrewheiss.com/example/cace/

Home Office (2025). Crime recording rules for frontline officers and staff, 2025/26. Home Office. https://assets.publishing.service.gov.uk/media/67ee9b2a199d1cd55b48c769/crime-recording-rules-for-frontline-officers-and-staff-2025_26-april-2025-update.pdf

Jolliffe, D., & Farrington, D. P. (2014). Self-Reported Offending: Reliability and Validity. In *Encyclopedia of Criminology and Criminal Justice* (pp. 4716–4723). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-5690-2_648

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kilkelly, P. U. (2023). *Evidence-based core messages for youth justice*. Justice Inspectorates. https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2023/10/Academic-Insights-Evidence-based-core-messages-for-youth-justice-by-Professor-Ursula-Kilkelly-Final.pdf

Klockars, A. J., & Sax, G. (1986). *Multiple Comparisons*. SAGE Publications. https://doi.org/10.4135/9781412985185

Krzywinski, M., & Altman, N. (2017). Classification and regression trees. *Nature Methods*, *14*(8), 757–758. https://doi.org/10.1038/nmeth.4370

McFadzien, K., & Phillips, J. M. (2021). Perils of the Subjective Approach: A Critical Analysis of the UK National Crime Recording Standards. *Policing: A Journal of Policy and Practice*, *15*(1), 556–569. https://doi.org/10.1093/police/paz023

Meurer, W. J., & Tolles, J. (2021). Interim Analyses During Group Sequential Clinical Trials. *JAMA*, *326*(15), 1524–1525. https://doi.org/10.1001/jama.2021.10174

Moore, S., & Pease, K. (2024). 'Focused Deterrence': Works in practice, but does it work in theory? *The Police Journal*, *97*(3), 541–558. https://doi.org/10.1177/0032258X231196116

Olphin, T., Bottomley, S., Webb, A., & Stock, M. (2024). *What Works Series: Focused Deterrence 12 Month Findings.* Thames Valley Violence Prevention Partnership. https://www.tvvpp.co.uk/wp-content/uploads/2024/10/What-Works-Focused-Deterence-FINAL-6.pdf

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Sage. https://us.sagepub.com/en-us/nam/realistic-evaluation/book205276

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Rioux, C., & Little, T. D. (2021). Missing data treatments in intervention studies: What was, what is, and what should be. *International Journal of Behavioral Development*, *45*(1), 51–58. https://doi.org/10.1177/0165025419880609

Rosenfeld, R., & Vaughn, P. (2024). *The Impact of Individualized Focused Deterrence on Criminal and Prosocial Outcomes*. CNA. https://www.ojp.gov/pdffiles1/nij/grants/309630.pdf

Senn, S. S. (2021). *Statistical Issues in Drug Development, 3rd Edition* (3rd ed.). Wiley. https://www.wiley.com/en-in/Statistical+Issues+in+Drug+Development%2C+3rd+Edition-p-9781119238577

Simanovic, T., McFarlane, P., Brennan, I., Graham, W., & Sutherland, A. (2024). Focused deterrence: A protocol for a realist multisite randomised controlled trial for evaluating a violence prevention intervention in the UK. *PLOS ONE*, *19*(3), e0301023. https://doi.org/10.1371/journal.pone.0301023

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Second Edition). SAGE Publications, Inc. https://uk.sagepub.com/en-gb/eur/multilevel-analysis/book234191

Stata Bookstore: Meta-Analysis in Stata: An Updated Collection from the Stata Journal, Second Edition. (n.d.). Retrieved 6 December 2024, from https://www.stata.com/bookstore/meta-analysis-in-stata/

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Sutherland, A. (2013). *A Methodology for Reconviction Studies Using Police National Computer (PNC) Data* (SSRN Scholarly Paper No. 2378345). Social Science Research Network. https://doi.org/10.2139/ssrn.2378345

Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M., & Rosenblum, M. (2020). *Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment* (No. arXiv:1910.13954). arXiv. https://doi.org/10.48550/arXiv.1910.13954

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817–838. https://doi.org/10.2307/1912934

Youth Endowment Fund. (2021). *Analysis Guidance*. Youth Endowment Fund. https://res.cloudinary.com/yef/images/v1623145483/cdn/6.-YEF-Analysis-Guidance/6.-YEF-Analysis-Guidance.pdf

Youth Endowment Fund. (2022). *Focused deterrence delivery framework*. Youth Endowment Fund.

Appendix 1: Power calculations for full trial period

```
## ACF1 Focused deterrence multicentre trial power simulations

# Required libraries
library(MASS)
library(tidyverse)
library(broom)
library(summarytools)
library(pscl)

# Constants for the simulation
num_sims <- 1000 # number of simulations to run per iteration
lambda <- 0.5 # mean for outcome distribution

# Define range of 'n' sample sizes and 'd' relative risk effect sizes to be tested
n_values <- seq(250, 3000, by = 250)  # set of sample sizes to test (increasing in multiples of 250)
d_values <- c(0.9, 0.8, 0.7, 0.6) # set of effect sizes to test

# Set seed
set.seed(123) # set seed to allow reproducible data

# Function to generate Poisson-distributed t1_outcome
generate_t1_outcome <- function(n) {
  rpois(n, lambda = lambda)
}


# Function to create risk variable correlated with t1_outcome
generate_risk <- function(t1_outcome, desired_correlation) {
  std_dev_noise <- sqrt(1 - desired_correlation^2)
  noise <- rnorm(length(t1_outcome), mean = 0, sd = std_dev_noise)  # random noise generation
  risk <- desired_correlation * t1_outcome + noise # generate risk score that is correlated with baseline outcome
  quantile_breaks <- quantile(risk, probs = c(0, 1/3, 2/3, 1)) # split simulated data into tertiles (low, medium and high)
  cut(risk, breaks = quantile_breaks, labels = c("low", "medium", "high"), include.lowest = TRUE) # categorise tertiles
}


# Function to create sample dataset
generate_sample_df <- function(n, d) {
```

```
  t1_outcome <- generate_t1_outcome(n)  # generate baseline outcomes
  risk <- generate_risk(t1_outcome, 0.6) # generate risk
  id <- seq_len(n) # generate unique ID
  condition <- sample(c(0, 1), replace = TRUE, size = n) # randomly allocation
  lambda_t2 <- ifelse(condition == 1, d * lambda, lambda) # create different means for
intervention and control that differ by effect size
  t2_outcome <- rpois(n, lambda = lambda_t2)
  tibble(id, condition, t1_outcome, risk, t2_outcome) # put columns together
}


# Function to create mean outcomes by allocation
run_model_with_means <- function(n, d) { # create function to run sample simulation with
varying t2 means
  sample_df <- generate_sample_df(n, d)
  sample_df %>%
    group_by(condition) %>%
    summarise(
      mean_t1_outcome = mean(t1_outcome), # generate mean baseline outcome
      mean_t2_outcome = mean(t2_outcome) # generate mean follow-up outcome
    )
}


# Function to run zero-inflated Poisson model
run_model <- function(n, d) {
  sample_df <- generate_sample_df(n, d) # generate simulated sample
  model <- zeroinfl(t2_outcome ~ condition + t1_outcome + risk, data = sample_df, dist =
"poisson") # run model
  coefs_df <- coef(summary(model)) # extract model summary
  coefs_df$count["condition", "Pr(>|z|)"] # return p-value for intervention covariate
}


# Function to run simulations and get means for t1_outcome and t2_outcome by condition
run_simulations_for_n_d <- function(n, d) {
  simulation_means <- map_dfr(1:num_sims, ~run_model_with_means(n, d), .id =
"simulation") # extract means
  mean_of_means <- simulation_means %>%
    group_by(condition) %>%
    summarise(
      overall_mean_t1_outcome = mean(mean_t1_outcome), # extract mean outcomes for
baseline by condition
      overall_mean_t2_outcome = mean(mean_t2_outcome) # extract mean outcomes for
follow-up by condition
```

```
  )

  simulation_results <- map_dbl(1:num_sims, ~run_model(n, d)) # run model and extract p-values
  power <- mean(simulation_results < 0.05, na.rm = TRUE) # power = proportion of p-values < 0.05

  list(mean_of_means = mean_of_means, power = power) # print power and means
}


# Run simulations for each combination of 'n' and 'd'
results <- list() # prepare list for results

for (n in n_values) {
  for (d in d_values) {
    cat("Running simulation for n =", n, "and d =", d, "\n") # Message: what combination is running now
    result <- run_simulations_for_n_d(n, d) # jumanji! run simulations
    results[[paste("n", n, "d", d, sep = "_")]] <- result # extract results
  }
}


# Save results
simulation_summary <- tibble( # create tibble of results
  n = rep(n_values, each = length(d_values)), # row of n
  d = rep(d_values, times = length(n_values)), # row of d
  power = map_dbl(results, "power") # populate tiblle with power results
)


# Generate line plot
ggplot(simulation_summary, aes(x = n, y = power, color = as.factor(d), group = d)) +  # create plot aesthetic n on a-axis, power on y-axis, grouped by d to create one line for each iteration of d
  geom_line(size = 1) + # generate line plot
  geom_point(size = 2) + # add data points for power corresponding to each iteration of n*d
  labs(
    title = "Statistical power across four effect sizes for primary outcome: number of VAP offences", # labels
    subtitle = paste("Simulated power for", num_sims, "simulations"),
    x = "Sample size",
    y = "Power",
    color = "Effect size"
```

```
  ) +
  geom_hline(yintercept = 0.8) + # add line to indicate threshold for 80% power
  theme_minimal() # make graph nice and clean

# Save the plot to file
ggsave("ACF1_FD_SAP_power_vs_n_for_d_simulations.png", width = 10, height = 6)

# Save simulation results to CSV
write.csv(simulation_summary, 'ACF1_FD_SAP_simulation_results.csv')



#code last run successfully by AS 29/11/2024
#annotated by IB 13/06/2025
```

Appendix 2: Sample size projections for full trial protocol

As of June 2024, the trial had a total sample size of 2,085. This is an addition of 2.25 cases per day across the combined trial for the period January to June 2024, which reflects the implementation of the mature intervention (see Table A1). At that rate, a further 780 individuals are projected to be identified and randomised in the remaining delivery period (with randomisation ending 120 days prior. This is projected to result in a final sample size of 2,865 individuals (see Table A2).

Table A1. Study throughput

Trial	Launch date	Days active	Cases to date	Interim period entry rate	Intervention cases
Coventry trial 1	06/06/2023	390	190	0.03	95
Coventry trial 2	14/07/2023	352	219	0.46	104
Leicester	22/05/2023	405	463	0.11	230
Manchester	20/07/2023	346	526	0.73	264
Nottingham	20/06/2023	376	298	0.38	148
Wolverhampton trial 1	06/06/2023	390	250	0.25	123
Wolverhampton trial 2	29/06/2023	367	139	0.29	62
Total		**2626**	**2085**		**1026**

Table A2. Project full trial sample sizes

Trial	Days remaining	Projected additional cases	Projected total cases
Coventry trial 1	307	9	199
Coventry trial 2	307	141	360
Leicester	307	34	497
Manchester	429	313	839
Nottingham	307	117	415
Wolverhampton trial 1	307	77	327
Wolverhampton trial 2	307	89	228
Total	**2271**	**780**	**2865**

Appendix 3: Updated CMO configurations

The Context-Mechanism-Outcome (CMO) configuration (see Figure A1) summarises the essential variations of programme delivery across different sites, recognising the diverse local contexts in which the intervention and its components are delivered.

Context

Qualitative data, observation and analysis of programme manuals (Brennan et al, 2024) have demonstrated that delivery models vary according to the delivery organisation, with the most significant difference being whether or not the intervention is police-led or not. At the contextual level, police acting as the referral and cohort identification vehicle, the intervention planner and the intervention deliverer (navigator) has resulted in differences in how the programme is conceptualised, led, governed, staffed, resourced and coordinated.

Mechanism

The contextual factors (police-led vs not police-led) have influenced the way in which deterrence activities and support mechanisms can be delivered and are likely to be received by cohort members. We observed differences in the coordination, quality and engagement of support services that are contributing to a more coherent delivery of the focused deterrence model (Brennan et al, 2024). For example, targeted deterrence and enforcement (mechanism 1) and individualised support (mechanism 2) are delivered differently if the individual or organisation delivering the message and offering or providing a person-centred service is a police officer or not. For example, police officers embodied a credible threat of consequence that a civilian navigator cannot. We have observed and recorded differences in how police and civilians introduce the programme and how they deliver the deterrence and the support messages in early interactions with cohort members. A police officer delivering the message may strengthen the deterrence mechanism but may simultaneously undermine the credibility and attractiveness of a person-centred support offer. In terms of community voice to influence and supporting desistance (mechanism 3), police are unlikely to embody a cohort member's community while a civilian navigator can. Furthermore, in cases where a navigator has lived experience of involvement in violence, they can exemplify the potential for desistance.

Outcomes

Being at the implementation stage of the intervention and having no prior comparison between police-led and not police-led interventions, we have no theoretical or empirical basis on which to differentiate between outcomes according to context and mechanism. Therefore, our outcomes - violence and co-offending - are homogeneous across all trials.

Figure A1. In-progress Context-Mechanism-Outcome framework (December 2024)

Contexts	Mechanisms	Proximal outcomes
	[Targeted enforcement m_1]	
Heterogeneity in resource availability, delivery models (e.g. police-led or not) and community-police trust: Local context affects how intervention is delivered and received.	**Soft and Hard enforcement**: Increased targeted deterrence activities, which range from soft enforcement (e.g., warnings, informal controls) to hard enforcement (e.g., tagging, control orders), impact participants' perceptions of certainty of arrest and punishment. **Mechanism interaction**: Effectiveness of deterrence is reinforced by strong community validation and reliable individualised support but weakened if community trust is low or support services are inconsistent.	**Progressive behavioural adjustment:** Participants may show partial engagement with non-violent modifications, relapse and eventually move toward sustained engagement with individualised support and compliance with norms.
	[Person-centred support m_2]	
Variance in resource allocation, coordination, engagement and spectrum of available support services: The level and quality of support services differ between sites. Some areas may have well-coordinated, resource-rich services, while others are more fragmented, affecting the effectiveness of some support mechanisms.	**Support variability**: The availability of individualised support packages varies in consistency and depth across sites. Navigators may provide stronger, more reliable (or credible) support in some areas, while in others, threats to withdraw support may be less believable, affecting engagement. **Mechanism interaction**: Engagement with support is most effective when reinforced by effective targeted deterrence and community engagement but weakened if the deterrence mechanism is inconsistent or community involvement is low.	**Variance in support engagement:** Engagement with support may fluctuate, with participants showing partial desistance, inconsistent engagement, or relapse before eventually adopting legitimate social norms.
	[Community Validation m_3]	
Decrease in levels of community confidence in local policing and statutory/non-statutory support services: Varying degrees of community trust and involvement influence the strength of community validation. Areas with higher community engagement are more likely to see stronger behavioural compliance and desistance outcomes.	**Levels of community engagement**: Community involvement exists on a spectrum (low, medium or high), with varying degrees of engagement affecting behavioural compliance and programme legitimacy. Strong community moral voice, supported by peer and familial influences, plays a key role in reinforcing prosocial behaviours. **Mechanism interaction**: Effective deterrence and support mechanisms strengthen community validation. If enforcement is weak or support inconsistent, the community's influence is undermined, leading to weaker compliance and outcomes.	**Community-driven compliance:** Community validation may lead to a gradual increase in legitimacy and compliance, with participants showing partial shifts in social capital and normative behaviour, moving toward desistance over time.

Appendix 4: Analysis R Code

Intention-to-treat

Intention-to-treat analysis code for primary outcome - count of violent offences.
Assuming that there is zero-inflation and that the outcome is over-dispersed.

```
library(glmmTMB)

model <- glmmTMB(t2_vap ~ condition + t1_vap + risk + ethnicity + sex + (1| trial), ziformula = ~1, family = nbinom2, data = acf1_df)
```

CACE R Code

In the event of very high levels of compliance the CACE and ITT estimates will be similar (with perfect compliance they would be identical). However, if we have non-compliance then CACE and ITT can diverge considerably.[8] We will also add in the same control variables as for the ITT analysis. Our definition of compliance is whether those allocated to the treatment group actually received intervention of any sort and whether those in control remained in control and did not receive the intervention and will be a binary variable.

```
library(AER)

model_iv <- ivreg(t2_vap ~ complier + t1_vap + risk + ethnicity + sex | condition + t1_outcome + risk + ethnicity + sex, data = acf1_df)
```

Co-offending

```
library(glmmTMB)

model <- glmmTMB(t2_coof ~ condition + t1_coof + risk + ethnicity + sex + (1| trial), ziformula = ~1, family = nbinom2, data = acf1_df)
```

Survival model R code

```
library(dplyr)
```

[8] https://www.rdocumentation.org/packages/estimatr/versions/1.0.4/topics/iv_robust

```
library(survival)

library(survminer)


# Mutating 'status' based on 't2_vap' in acf1_df

acf1_df <- acf1_df %>% mutate(status = as.numeric(case_when(t2_vap == 0 ~ 0, t2_vap > 0 ~ 1, TRUE ~ NA)))


# Cox model

cox_model_fixed <- coxph(Surv(time_to_offence, status) ~ condition + t1_vap + ethnicity + sex + risk + trial, data = acf1_df)
```

Meta analysis approach

We are primarily using the meta-analysis approach to illustrate between site / between group variation in treatment effects. We will do this on the basis of unadjusted data (unadjusted for baseline/ covariates), which will consist of:

- Site location/trial.
- Sample size for control & treatment groups based on those randomised.
- Outcomes in control and treatment:
    - the number of those who reoffended for basic prevalence.
    - the count of offences for those who have reoffended.
- Subgroup variables (as detailed above).

That data will be set out in a spreadsheet, inputted into the Comprehensive Meta-Analysis (CMA) programme (Borenstein et al., 2022) and outputted as forest plots that illustrate the between-group variation. As above, the intention-to-treatment (ITT) statistical model based on the PNC data is the definitive result for this trial.

Appendix 5 Interim analysis in early implementation report

Figure A1 below is the CONSORT diagram for the interim analysis, which covered the period from intervention initiation (May/June 2023) to December 2023.

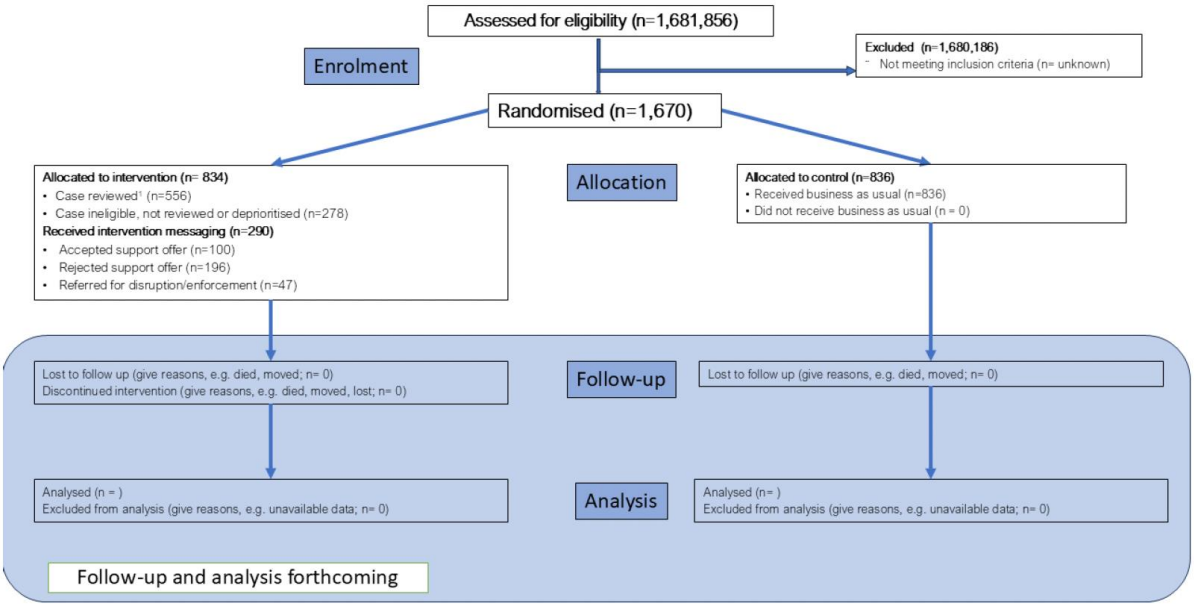


Figure A1. Interim analysis CONSORT diagram

As part of our reporting to YEF we undertook preliminary analysis based on early implementation data as part of our stopping rules assessment, following the approach set out in the protocol for the analysis model and reporting of the interim analysis (i.e. not reporting p-values). In addition, we also reported interim results as blinded when presenting in meetings with sites and funders. That is, we reported group means and differences but masked which group was which (and did not give that detail in reports to YEF).

The early implementation analysis was useful in several ways. First, it allowed us to test our analysis approach in a low-stakes way. Second, it enabled us to talk to stakeholders about the focus of the interim analysis on informing decisions around trial progress / backfires in a workshop held in Manchester early in 2024. Third, it illustrated the need to incorporate time since randomisation in our analysis model (set out above). This is because in some instances actual intervention contact did not begin until several weeks or months later in some instances, owing to intentional design (in West Midlands) or capacity constraints (in Nottingham and Manchester). This means there is a period of time between randomisation

and intervention that varies between trial and individuals within trial - adding variance and reducing power (Note that anchoring the analysis around the point that intervention delivery actually began would introduce substantial bias into estimates because when intervention begins is non-random and suffers from selection bias, whereas entry into treatment or control *is* random.) Finally, the pilot analysis illustrated that trial need to ensure that they reduce their 'backlog' of cases before assessing others for randomisation because the initial 'batch' of randomisations for all trial led to over-supply of eligible cases versus capacity to work with them. The evaluation team put this to sites at the one-day workshop and it was agreed to change the referral and randomisation approach.

Interim analysis of quantitative data

To conduct our interim analysis of outcomes after up to six months (Brennan et al., 2024; using data up to December 2023), data were obtained from site analysts on the prevalence and frequency of two outcomes (violence against the person offences; co-offending offences). 'Up to six months' is important to note because it means those analysed have different 'time at risk' - specifically some might have been randomised the day before we obtained data, and some might have been randomised six months ago. We took the decision to analyse everyone because to only include those with six months of data would have meant a substantially smaller sample size. It should also be noted that the differential time at risk is equivalised across treatment and control.

As noted above, consistent with best practice, we have blinded the treatment condition in the data to reduce the risk of premature conclusions being made about the direction or intensity of treatment effects. It should also be noted that these data were provided in aggregated form, which prevents statistical testing of differences between the groups and also prevents the calculation of 95% confidence intervals or calculation of the relationship between the outcome and any element of the intervention.

The results present a somewhat complicated picture of the outcome at the end of the early implementation phase. Between the groups, the relative risk of being involved in violence was 0.85. Conversely, however, the relative reduction in the frequency of violent offences between the same groups was 0.92 (1/1.08) in the opposite direction. In simple terms, individuals in condition A were less likely to be involved in violence, but when they were, it was more frequent. The co-offending outcome presents a similar contradicting story. Individuals in condition A were less likely to be involved in co-offending, but when they were, they did so more frequently than individuals in condition B.

Table A3. Distribution of outcomes from early implementation (blinded condition)

	Condition A	Condition B	Relative risk
Violence against the person offences			
Prevalence of violence against the person	0.17	0.20	0.85
Rate of violence against the person offences per person	0.38	0.36	1.08
Co-offending offences			
Prevalence of co-offending	0.15	0.15	0.96
Rate of co-offending per person	0.35	0.29	1.20

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413