EVALUATION PROTOCOL

The London Young People Study: An evaluation of the Your Choice intervention

Institute of Fiscal Studies (IFS) and Anna Freud Centre (AFC)

YOUTH ENDOWMENT FUND

The London Young People Study: An evaluation of the Your Choice intervention

Evaluation protocol

Evaluating institution:

Principal investigator(s):

YEF trial protocol for efficacy and effectiveness studies

Project title[1]	The London Young People Study: Evaluation of the Your Choice Programme Using a Cluster Randomised Control Trial
Developer (Institution)	London Innovation and Improvement Alliance (LIIA) Violence Reduction Unit (VRU)
Evaluator (Institution)	Institute of Fiscal Studies (IFS) and Anna Freud Centre (AFC)
Principal investigator(s)	
Protocol author(s)	
Trial design	Two-armed cluster randomised controlled trial with random allocation at the youth practitioner level
Trial type	Efficacy
Evaluation setting	28 Local Authorities of London
Target group	Any child aged between 11-18 years old who is assessed as medium or high risk of harm / vulnerability as a result of extra-

[1] Please make sure the title matches that in the header and that it is identified as a randomised trial as per the CONSORT requirements (CONSORT 1a).

	familial harm and has been considered by a multi-agency panel (typically MACE / Pre-MACE)
Number of participants	Between 1700 and 1800 young people across the internal pilot and efficacy stage Across teams that have ever been randomised, this includes an expected 1698 young people recruited between August 2023 and December 2024 and 83 young people who were recruited during the internal pilot phase and who have endline data. We expect a 70% data completion rate, which sets the overall sample for impact evaluation at 1,272.
Primary outcome and data source	Indicator for scoring in the high or very high range of the conduct problems subscale of the Strengths and Difficulties Questionnaires (SDQ). The high and very high range threshold is defined by the SDQ's four-fold categorisation. The SDQ is taken at the start and end of the young person's pathway through the programme, at weeks 1 and 20 after recruitment. In both cases, they are administered during a session with the practitioner, as part of the larger baseline and endline surveys. Some small variation in the timing of these surveys is allowed, to accommodate differing start and end dates for the programme and convenience in scheduling meetings between participants and practitioners. Below we will be referring to week 1 and 20 as expected times for these surveys.
Secondary outcome and data source	*Criminal activity*: Indicator for recorded arrest in Police National Computers during first 16 months after recruitment (allowing for 4-months treatment plus 1 year following treatment). This outcome describes comparatively serious offences, which are more likely associated with violent behaviour. *Self-reported and practitioner-reported perceptions of young person's safety*: Young person and practitioner versions of "Checkpoint. A safety scale for young people", which is an instrument to measure young people's perceptions of safety

	developed by the research team. Measured at baseline and endline as part of the participant's and practitioner's surveys (weeks 1 and 20 after recruitment, respectively). *Wellbeing*: The Short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS). Measured at endline as part of the participant's survey administered in session with practitioner at week 20 after recruitment. *Emotional self-regulation*: Trait Emotional Intelligence Questionnaire – Adolescent Short Form (TEIQUE-ASF) – Self regulation subscale. Measured at endline as part of the participant's survey administered in session with practitioner at week 20 after recruitment. *Social connectedness*: Social Connectedness Scale – Revised (SCS-R). Measured at baseline and endline as part of the participant's surveys administered at weeks 1 and 20 after recruitment. *Internalising behaviours*: emotional difficulties and peer difficulties subscales of the SDQ measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Hyperactivity*: hyperactivity subscale of the SDQ measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Prosocial behaviours*: Strength and Difficulties prosocial behaviour subscale measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Prosocial identity*: Pro-social Identity Scale (PIDS) measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively.

Protocol version history

Version	Date	Reason for revision
1.2 [*latest*]		
1.1		
1.0 [*original*]		*[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).

Table of contents

Study rationale and background

Theoretical and scientific background, policy and practice context and rationale for the Your Choice intervention[2]

The Your Choice programme is a Cognitive Behavioural Therapy (CBT) enhanced approach to practice, delivered through high intensity contact within adolescent services. The 12-18-week programme is delivered by specially trained practitioners, who are trained in CBT tools and techniques and are supported by regular clinical supervision. Training for practitioners is delivered through a train the trainer model by clinicians with experience in the delivery of CBT.

Background

There is growing evidence that therapeutic support for unmet needs, adverse or traumatic experiences, and other risk factors may prevent young people from becoming involved in crime and violence or may reduce further involvement (Gaffney et al., 2021). However, young people with the highest levels of risk factors are currently least likely to access such support in a clinical setting (Crenna-Jennings & Hutchinson, 2020). Lack of availability, poor information, inflexibility, complicated referral processes, cultural barriers and stigma impact on young people's access to clinical services (Brooks et al., 2021). Yet, due to concerns regarding risk, harm and/ or vulnerability these young people are likely to be accessing support from other adolescent support or statutory agencies. The Your Choice programme seeks to upskill those practitioners in a range of CBT tools and techniques that they can weave in to their existing practice frameworks and work with young people.

Young people who meet the threshold for adolescent services are likely to have or be experiencing Adverse Childhood Experiences (ACE's) and/or childhood trauma. Neurodivergence is also common. Consequently, this cohort are more likely to find it difficult to recognise and manage different emotions and behaviour and are at increased risk of presenting in distress and developing mental health difficulties. There is also an increased propensity for "risky" behaviour (Sweeney, Clement, Filson & Kennedy, 2015).

Cognitive Behavioural Therapy (CBT)

In the UK, the National Institute for Health and Clinical Excellence (NICE) recommends CBT for the management of many common mental health difficulties. This is because it has been tried

[2] See Your Choice page on the London Innovation and Improvement Alliance (LIIA) website for further information on the background of Your Choice, including references.

and tested with various populations for dealing with a wide range of psychological "problems or challenges" (Butler, Chapman, Forman & Beck, 2006).

A key concept of CBT is that an individual's thoughts, feelings, bodily sensations and behaviour are interlinked. The 5-factor model in CBT demonstrates the connections between situation, thoughts, emotions, bodily sensations and behaviours in the context of internal or external triggers (the situation/ environment). These 5 factors are considered to be so closely connected that changes in any one of these can lead to changes in the other factors.

CBT describes a family of interventions. In more traditional (first and second wave) CBT, thinking and behaviour are key targets for change as it is thought that these factors can be most readily influenced as an individual has more control over those parts of the system. However, rather than changing the content of an individuals' thoughts and inner experiences, newer third wave cognitive behavioural therapies take a broader approach and seek to bring an awareness and different relationship that an individual has with their inner world (Eels, 1997). The Your Choice programme includes tools and techniques from traditional CBT (i.e. thought challenging) and third waves approaches (i.e. emotion recognition and regulation).

What is Your Choice and how is it adapted?

Despite the promising evidence base, a common criticism of CBT has been an over reliance on the mechanistic application of a set of techniques, with a lack of emphasis of the importance of the therapeutic relationship (Wenzel, 2021). Consequently, the Your Choice programme prioritises time for investing in and nurturing relationships with young people through intensive contact. This is in acknowledgement of the needs of the cohort and that relationships that are safe, collaborative and trusting are likely to impact on engagement and outcomes (Thomson et al, 2007).

In addition to being informed by the principles and practices of CBT, the Your Choice programme is underpinned by a range of psychological theory and best practice principles relevant to working with adolescents at risk. This includes attachment and developmental theory and child first and trauma informed principles.

In accordance with child first principles, the Good Lives (Ward & Steward, 2003, cited in Barnao, 2022) and public health approach (Case and Browning, 2021) to reducing risk, Your Choice is a strengths-based programme, which is responsive to an individual's interests, values and aspirations and builds capabilities and strengths in order to reduce risks. Your Choice emphasises the importance of behavioural activation (a core component of CBT) to promote positive and pro social behaviour by i) providing access to positive alternatives and opportunities, ii) connecting young people with "safe" communities iii) disrupting and diverting young people away from unhelpful, unhealthy patterns of behaviour iv) empowering young people to identify with "routes out" of problematic or harmful behaviours

and/or situations. This is particularly important given that adolescence is a significant developmental phase relating to the formation of identity (Erikson, 1968).

Advances in brain imagery techniques have developed the understanding of the significant neurological development that occurs during adolescence. This includes a significant re-organisation of the brain as it undergoes intense synaptic pruning, where neurological pathways that are used are re-enforced and those that are not are pruned away. We now know that the brain matures in back to front order with the control centre (the prefrontal cortex) the last part of the brain to develop. All the while the reward centre is hypersensitive, which helps to understand why there is a spike in risk taking behaviour during adolescence (Steinberg, 2008). The high intensity contact provided through the Your Choice programme encourages young people to engage in structured and appropriate reward seeking behaviour alongside the mature frontal cortexes of their Your Choice trained practitioners.

Your Choice is underpinned by trauma informed principles, as evident in the infrastructure, design and content of the programme. This is in recognition of the high prevalence of exposure to trauma for both the young people and the families that are supported through the programme but also the practitioners who are likely to be exposed to vicarious trauma through their work. Within the programme there are practical tools to improve emotional literacy to support the recognition and management of intense emotions, which may be consequent of traumatic exposures and experiences.

Emerging evidence relating to the short- and long-term impact of exposure and repeated exposure to vicarious trauma can inhibit practitioner's ability to provide high quality care (Quitangon, 2019). Therefore, regular access to clinical supervision is a core requirement of Your Choice programme delivery to ensure that practitioners have an opportunity to consider and work through the impact of exposure to trauma through their work.

Overview of the evaluation design

The London Young People Study (LYPS) is a mixed method evaluation of the Your Choice intervention.

The impact evaluation is based on a two-armed cluster randomised controlled trial (RCT) where the unit of randomisation is teams of youth practitioners. That is, teams supporting young people eligible for Your Choice are randomly assigned to train in and deliver Your Choice (treatment group) or to supporting young people following Business As Usual (BAU) practices (control group). Teams randomised out of training during the efficacy trial will be offered to be trained in Your Choice later on. This was explicitly agreed with LAs given the reservations they had about randomisation. It seems unlikely that control teams react to the possibility of future training so much in advance. LAs also have an obligation towards the young people they see, to deliver statutory services according to the BAU processes. For these reasons, this setting is unlikely to affect results.

The RCT design does not require that young people are randomly assigned to teams, as individual level randomisation of young people would interfere with the usual delivery of statutory services in ways that would be impractical for and unacceptable to Local Authorities. A key aspect of the Local Authorities' objection to the randomisation of young people was that they could not guarantee the availability of allocated spaces given the limited capacity of teams and how that capacity fluctuates over time, rendering the process difficult to manage and likely to create serious disruptions in services. Instead, the proposed RCT design is drawn under the expectation that the assignment of young people to teams is based on team availability at the time of referral, in a system that works at capacity. The one exception to this rule is for young people returning to LA services after a brief interruption, who would be assigned to the same team that previously worked with them. Such an assignment rule, if rigorously followed, effectively guarantees that assignment is independent of the team status regarding Your Choice training. Evidence collected during the pilot trial showed that the observed characteristics of young people assigned to treated and control teams were well balanced, and that capacity and historical assignment were major drivers of assignment of young people to teams, as described in the next point below.

The implementation and process evaluation uses both quantitative and qualitative methods, including collection of process data on delivery of Your Choice and BAU support as well as qualitative interviews of LA staff and young people involved in the delivery of Your Choice.

Learnings from the pilot trial of Your Choice and how they feed into the efficacy trial design

The implementation and delivery of Your Choice was piloted in the context of youth services delivered by London Local Authorities during the second half of 2022 and early 2023. The aim of the pilot study was to train and initiate youth services in Local Authorities in the delivery of Your Choice in an experimental setting to inform the feasibility and practical aspects of delivering the intervention and implementing a successful efficacy trial.
Specifically, the objectives of the pilot study were:

1. To assess the feasibility of implementing an effective data collection exercise that supports the quantitative evaluation of Your Choice;
2. To examine how the Your Choice intervention is implemented, fidelity of delivery, and what helps and hinders implementation;
3. To assess the adherence of Local Authorities and youth practitioners to randomisation;
4. To pilot study outcomes and evaluation methods, assess the parameters for conducting an efficacy evaluation and to assess whether operational progression criteria have been met and if so to develop a full protocol for an appropriately powered efficacy study.

Detailed findings from the pilot study can be found in the Your Choice Interim Pilot Report. Crucial to the Efficacy Trial Protocol, it was found that:

- In the majority of cases, local councils adhered well to the randomisation of teams to training.
- Evidence on the allocation of young people to trained and untrained teams indicates that the distribution of observed characteristics at baseline are balanced across treatment arms. Moreover, qualitative information showed that the capacity of teams was an important driver of the allocation of young people to teams, a criterion that is independent of the young person specific characteristics and potential benefit from participating in Your Choice. This evidence supports the assumption that the allocation of young people to teams was as good as random, which is required for the validity of the evaluation exercise.

  However, qualitative data also revealed that some allocators considered potential benefits and perceived level of risk in deciding how to allocate young people to teams, a practice that violates the principles of the evaluation design (which is based on the assumption that assignment to teams is based on team capacity at the time the young person is referred to the services, with one exception for young people who are returning to LA services after a brief interruption, and who are assigned to the same team they were previously). The consequences of considering potential benefits from Your Choice in determining assignment are not evident from observed pilot data on baseline characteristics (as far as we can say, observed characteristics are well balanced across treatment arms), but the suggestive evidence of its (however sparse) use prompted us to reinforce the training of Local Authorities staff ahead of efficacy around the efficacy trial and especially the importance of rigorously following the RCT design at all stages of delivering Your Choice for the purpose of the evaluation.
- Session forms and qualitative data reveal that the intervention was implemented with fidelity, with those assigned to teams delivering Your Choice receiving a more intense schedule of meetings that were more geared towards CBT methods and contained different content as compared to BAU.
- The recruitment of young people into the pilot study was lower than anticipated. This affects our expectations about the size of the experimental group during the efficacy trial, and is reflected in the power calculations detailed later in this report.
- The sharing of data between councils and evaluators was carried out at irregular time intervals and sometimes with severe delays; this was particularly the case with process data collected through session forms (required for the implementation and process evaluation). The collection of endline survey information, both from young people and practitioners, was also patchy and often delayed. These concerns about the data prompted us to develop a new online mobile-optimised interface for practitioners to use to fill out forms in a much easier and faster way than during the pilot and for SPOCs to supervise data collection in real-time. Moreover, to improve the timely collection of

data at endline in the efficacy study, which is critical for measuring the primary outcome, the endline youth survey started being administered in a regular session at the end of the programme, in the presence of the youth practitioner and with an audio facility for participants who have difficulties reading the text. This changes the practice briefly tried out at the start of the pilot period, when young people would be contacted outside regular sessions to complete the survey with the support of a peer researcher. The majority of pilot participants already filled in their endline questionnaire using the new procedure.

- Obtaining the collaboration of control teams was especially difficult given the generalised feeling among youth practitioners in these teams that there is less in the trial for them and the young people they treat. It led to lower adherence to data collection protocols among control teams. In response, the training campaign prior to efficacy trial will aim to reach SPOCs and all practitioners and emphasize the critical role of control teams for the success of the efficacy trial and future support for expansion of Your Choice in the longer term.
- Early testing suggests that the SDQ information collected at baseline is of good quality, supporting the use of a similar test at endline and for the primary outcome.
- The intervention originally projected that the delivery of Your Choice to young people would immediately follow recruitment, running between weeks 1 and 12 and delivering 3 weekly sessions over this period (calls, meetings, going to the gym, working with parent/carer for psychoeducation), each session lasting for 45-60 mins (possibly longer). The pilot demonstrated that this time schedule is difficult to keep in practice. More realistically, there is generally a gap of up to 2 weeks between recruitment and the starting of the programme, and the delivery of sessions happens over a longer time frame of 16 to 18 weeks. The programme is typically finished by the end of week 20. For that reason, the efficacy trial and the data collection exercises are designed under the assumption that the delivery of Your Choice to young people takes 20 weeks from recruitment to end. We refer to the first and last weeks in the programme as weeks 1 and 20, respectively.

Intervention

- [34]The intervention description using the TIDieR Framework (Hoffman et al. 2014), reproduced below.

Item	Description
BRIEF NAME	Your Choice
WHY	Young people who get involved in violence (those most at risk) are those in most of need of therapeutic support, but most unlikely to receive it. We need to shift how we offer support to young people, by shifting the offer, so they can access it within their community, within a broader context of support and behavioural change. This can be delivered best through a holistic, community model delivered through all relevant partners.
WHAT	Your Choice is for any child aged between 11-18 years old who is assessed as medium or high risk of harm / vulnerability as a result of extra-familial harm and has been considered by a multi-agency panel (typically MACE / Pre-MACE). Your Choice includes three main components: 1. Upskilling practitioners via 5 days of training for youth workers (delivered in a cascading model) and the provision of handbook and resources to supporting delivering training sessions 2. Intensive work with children to build an authentic and trusting relationship with the practitioner and create a safe space where young people can grow, understand and formulate their needs and goals. Specifically:

[5] At the time of writing, 4 Local Authorities had not confirmed their expected numbers of participants and we therefore made conservative assumptions about their expected numbers of participants.

[5] At the time of writing, 4 Local Authorities had not confirmed their expected numbers of participants and we therefore made conservative assumptions about their expected numbers of participants.

	- Young people in the treated arm will receive the equivalent of 3 x weekly meetings with trained youth practitioner for 12 weeks, over the 20 weeks after recruitment. - The sessions will deliver an accessible clinical intervention, focusing on emotional literacy, emotion regulation, understanding cognitive processes, and strategies for managing intense feelings - Sessions will be informed by CBT tools and techniques, such as goal setting (using Goal Based Outcome Tool) and practical support with activities to achieve these goals 3. Monthly clinical supervision by clinical leads hired by Local Authorities
WHO PROVIDED	Youth practitioners: youth workers, social workers, youth justice worker, gang workers, hired by the Local Authorities as practitioners in the teams involved in the study
HOW	Individual or work with the family
WHERE	Range of locations, accessible to the young person, so they are engaged in the places they want to be engaged; mainly community settings such as youth centre, cafes, gyms, etc
WHEN and HOW MUCH	The intervention originally projected the delivery of 3 weekly sessions for 12 consecutive weeks (calls, meetings, going to the gym, working with parent/carer for psychoeducation); 45-60 mins per session (poss. Longer). The pilot demonstrated that this intensity was difficult to achieve. More realistically, the same total number of sessions are delivered over an extended period of 20 weeks from recruitment.
TAILORING	To facilitate sustainability and meet local needs, it is important that Local Authorities own Your Choice; it

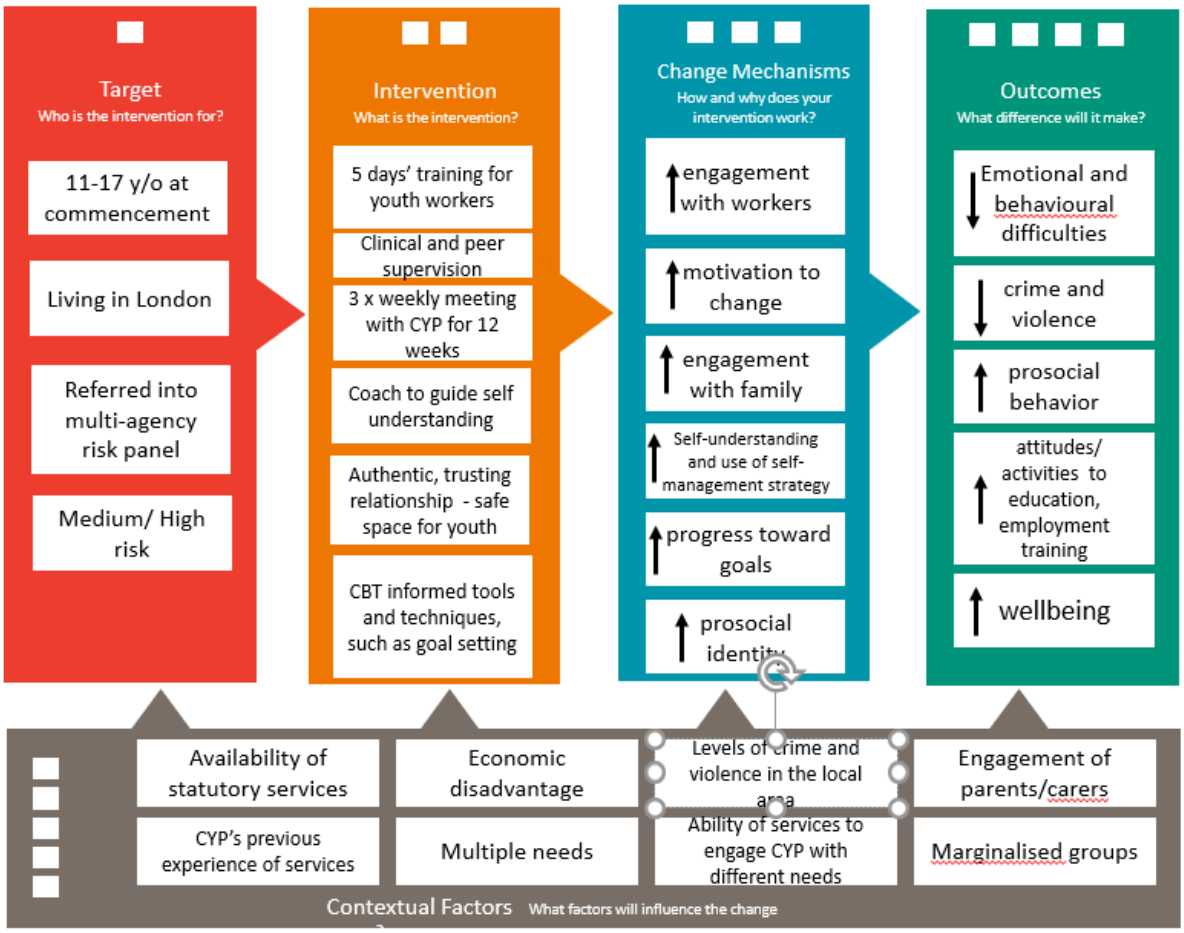
	will build on existing services and delivery for this cohort of young people, which will vary between different Local Authorities
MODIFICATIONS	Findings from the pilot have prompted some adaptations to the evaluation design (more details above, in the subsection 'Learnings from the pilot trial of Your Choice'): - The age range was extended to include 18 year olds - The endline youth survey will be administered in a regular session, in the presence of the youth practitioner, to avoid long delays in data collection and improve response rates; - The effective delivery period of Your Choice to young people was extended from 12 to 18 weeks in view of the practical limitations of concentrating the treatment in the shorter period. - Failure by a few allocators to follow exclusively the capacity and historical assignment rules in assigning young people to teams lead to the launch of a large training campaign ahead of the efficacy trial, to remind all those involved about the assignment, delivery and data collection protocols and the importance of following them strictly.
HOW WELL	To test fidelity monitoring during the efficacy trial, we will use qualitative interviews of practitioners and young people as well as analysis process data on the content of each session and the occurrence of clinical supervision sessions.

Further information about delivering the intervention can be found on the Your Choice page of the LIIA website (https://www.liia.london/your-choice). From there, practitioners trained in Your Choice can also access a secured portal providing them with even further tools to deliver Your Choice. The training of trainers is provided by Dr Karla Goodman, Practice Lead

at LIAA and developer of the programme. Trainers are for the most part clinical leads and team managers.

Logic model

The intervention is hypothesized to work through a number of mechanisms, described in the logic model below. The logic model is the same as the logic model we designed before the pilot trial, with the exception of one of the mechanisms, which was reformulated from increased help-seeking behaviour to increased self-understanding and use of self-management strategy. This change was informed by pilot qualitative findings.



Control condition

The control condition is the Business As Usual support that young people eligible for Your Choice would receive in absence of the study. BAU will vary between Local Authorities and range in terms of intensity and techniques practitioners use.

Incentives or restrictions for those in each group

There are no incentives or restrictions on the delivery of the Your Choice intervention among young people allocated to a trained team. Study participants will receive £10 voucher upon completion of the baseline young person questionnaire and £25 upon completion of the endline young person questionnaire.

Period when the intervention is being delivered

Delivery of Your Choice was expanded in three tranches. Delivery starts with the training of teams in CBT-informed tools and techniques, which took place during three periods:

- Home Office stage: The first training phase started in December 2021 and saw a small number of teams across 32 Local Authorities (LAs) in London trained in and deliver Your Choice. In 5 of these LAs, the trained team was randomly selected from two candidates (in all of the other 27 LAs, the trained teams were non-randomly selected).
- YEF (internal) pilot stage: Training was extended to additional teams during the spring of 2022, in preparation for the YEF pilot trial of Your Choice between August 2022 and March 2023. In the majority of (but not all) cases, the selection of treated and control teams was randomised within LA. In some LAs however such randomisation was not possible because they could not name two untrained teams to participate in the pilot.
- YEF efficacy stage: LAs in London have put forward 2 or more untrained teams for randomisation at this stage. The delivery of training to randomly selected teams amongst those will happened during the summer of 2023 in preparation for the efficacy trial.

Following the training of teams, the assignment of young people to treatment and control teams and the delivery of Your Choice has expanded accordingly, starting in January 2022 and will continue uninterruptedly but with increasing number of teams until the projected end of the efficacy trial, in December 2024.

How the intervention providers are assigned to groups

The treated and control groups are defined at the team level, based on whether or not a team was assigned to be trained in Your Choice. Once a team has been trained at any of the three training stages, it can only act as a treated team in subsequent phases. If a team acts as a control team in one stage, it can be re-randomised in the next stage and possibly become a treated team.

Impact evaluation

Research questions or study objectives

The research questions have been informed by the logic model and designed to reflect the complex set of relationships between the Your Choice intervention and the primary and

secondary outcomes, the mechanisms underlying these relationships and the importance of contextualising factors in determining the strength of these relationships. Some of the mechanisms will be explored via quantitative analysis and some will be explored via qualitative analysis. The data collection plan will allow us to explore some contextual factors and how these influence the change, but not all of them.

The research questions are as follows:

1. Can an intervention informed by CBT techniques and practices, and delivered by trained frontline practitioners, reduce conduct problems among young people most at risk of being affected by violence?

This is the main research question, focusing on the impact of Your Choice on the primary outcome – an indicator for scoring in the high and very high range in the conduct problems subscale from the Strengths and Difficulties Questionnaires (SDQ).

The SDQ conduct problems subscale is a measure of children's ability to behave in a socially acceptable manner, considerate of others. Scores in the abnormal range of this subscale are highly prevalent among the target population, as shown in the Your Choice Internal Pilot Report. They have also been related to aggressive behaviour and predisposition to rule breaking (Essau et al. 2012, Yao et al. 2008), and are likely an important predictor and precursor of disruptive and criminal behaviour (Spaan et al. 2023).

This combination of factors explains our choice of primary outcome. Specifically, we expect that successful violence reduction interventions are especially effective among individuals scoring in the abnormal range of the SDQ conduct problems subscale, and that conduct problems are at the forefront of the causal chain of effects that may ultimately see violence reduction result from the intervention. Focusing on the most relevant part of the scale seems most adequate for the purpose of detecting changes in predisposition to aggressive and violent behaviour. The alternative of considering the continuous conduct difficulties index runs into the risk of diluting effects by averaging over heterogeneous groups by conduct difficulties. Other scales in the SDQ, such as that measuring prosocial behaviour, though certainly relevant to provide a comprehensive characterisation of the effects Your Choice and to explore its potential mechanisms, are not as directly and closely related to the ultimate goal of violence reduction, and hence will be considered secondary outcomes.

Our primary outcome will be measured at endline in the participant's survey administered at week 20 after recruitment.

The null hypothesis we will be that there is no difference in this outcome between young people allocated to Your Choice and young people allocated to business as usual.

2. Is there a difference in offending behaviour between young people allocated to work with a team of practitioner trained in Your Choice in comparison to young people allocated to work with a team of practitioner delivering BAU support?

The outcome of interest is whether participants are ever arrested in the first 16 months after recruitment, as recorded in the PCN database. This is a key secondary outcome given that it directly relates to and informs on Your Choice's goal of reducing violence among young people at most risk. However, the low frequency of criminal activity will raise challenges in detecting impacts on this variable, and the time lag to obtaining crime data will necessarily delay the analysis. For these reasons, we consider the study of impacts on crime activity to be exploratory.

Two intermediate outcomes will also be studied: the self-assessed and practitioner-assessed level of risk, measured at endline as part of the young person and practitioner survey taken in week 20 after recruitment. These will provide and earlier indication of effects on self and practitioners' perceptions of likelihood of engaging in risky behaviour.

The null hypothesis we will be testing are: (i) that there is no difference in arrest probabilities between young people allocated to Your Choice and young people allocated to business as usual and (ii) that there is no difference in the self and practitioner assessments of the young person's likelihood to engage in risky behaviour, between young people allocated to Your Choice and young people allocated to business as usual.

3. What are the mechanisms through which Your Choice works? Specifically, in comparison to young people allocated to BAU support, do young people allocated to work with a team of practitioners trained in Your Choice:

- Gain higher wellbeing and emotional self-regulation? These outcomes will be measured as part of the endline young person survey completed in week 20 after recruitment, using the Warwick-Edinburgh Mental Wellbeing Scale, the internalising subscales of the SDQ, and the Self-regulation subscale from the Trait Emotional Intelligence Questionnaire – Adolescent Short Form (TEIQUE-ASF).
- Have stronger relationships with their practitioners, with their family and other sources of social support? The outcome of interest will be the Social Connectedness Scale – Revised (SCS-R), included in the endline young person questionnaire (taken 20 weeks after recruitment).

  Have a more prosocial behaviour? The outcome of interest will be the SDQ prosocial behaviour scale, measured as part of the endline young person questionnaire in week 20 after recruitment.
- Have a stronger prosocial identity? The outcome of interest will be the ProSocial Identity Scale, measured as part of the endline young person questionnaire

4. Do the impacts of being allocated to a team trained in Your Choice differ by gender, age, ethnicity, special education needs, level of risk assessed by the practitioner?

Design

The full study is composed of an internal pilot and an efficacy phase. The pilot started in August 2022, with teams randomly assigned to Your Choice training in December 2021 and later in the Spring 2022. Young people assigned to the treatment and control arms during the pilot period were provided with the exact same treatment conditions that are planned for the efficacy phase, which starts one year later in August 2023 with additional teams randomly assigned to training in the Spring and Summer of 2023. The assignment mechanism of young people to teams also follows the same rule across the two periods. The data collected for participants in both periods is also similar, and includes background data, needs presented, session data and baseline and endline surveys. In particular, all primary and secondary outcomes have been collected for participants during the pilot period, as will be for the efficacy trial and following the same procedures. The consistency in design across the two periods allows us to treat the pilot as internal. Details of the design can be found in Table 1 below.

Table 1: Trial design

Trial design, including number of arms		**Two-arm, cluster randomised**
Unit of randomisation		Teams of youth practitioners
Stratification variables (if applicable)		Local Authority
Primary outcome	variable	Conduct problems
	measure (instrument, scale, source)	An indicator for scoring in the high and very high range of the conduct problems subscale of the Strengths and Difficulties Questionnaires. Measured as part of the endline young person survey, administered at week 20 after recruitment during a session with the practitioner.

	variable(s)	Offending activity, mental wellbeing, emotional self-regulation, social connectedness, internalising behaviours, hyperactivity, self-reported and practitioner-reported perceptions of young person's safety, prosocial behaviour and prosocial identity.
Secondary outcome(s)	measure(s) (instrument, scale, source)	*Criminal activity*: recorded arrest in Police National Computers during the period of 16 months after recruitment. *Self-reported and practitioner-reported perceptions of young person's safety*: Young person and practitioner versions of "Checkpoint. A safety scale for young people", which is an instrument to measure young people's perceptions of safety developed by the research team. Part of the endline young person questionnaire, administered at week 20 after recruitment. *Wellbeing*: The Short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS). Part of the endline young person questionnaire, which is administered at some point between weeks 14 and 20 after recruitment. *Emotional self-regulation*: Trait Emotional Intelligence Questionnaire – Adolescent Short Form (TEIQUE-ASF) – Self regulation subscale. Part of the endline young person questionnaire, administered at week 20 after recruitment. *Social connectedness*: Social Connectedness Scale – Revised (SCS-R). Part of the endline young person questionnaire, administered at week 20 after recruitment. *Internalising behaviours*: emotional difficulties and peer difficulties subscales of the SDQ measured at baseline and endline as part of the participant's

		surveys administered following recruitment and at week 20 after recruitment respectively. *Hyperactivity*: hyperactivity subscale of the SDQ measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Prosocial behaviours*: Strength and Difficulties prosocial behaviour subscale measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Prosocial identity*: Pro-social Identity Scale (PIDS) measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively.
Baseline for primary outcome	variable	Conduct problems
	measure (instrument, scale, source)	Indicator for scoring in the high or very high range of conduct problems scale from the Strengths and Difficulties Questionnaires. Measured as part of the baseline participant's survey, administered in week 1 after recruitment, before the start of Your Choice.
Baseline for secondary outcomes	variable	Social connectedness score, internalising behaviour, hyperactivity, prosocial behaviour and prosocial identity, self-reported and practitioner-reported perceptions of young person's safety,
	measure (instrument, scale, source)	*Social connectedness*: The Social Connectedness Scale – Revised (SCS-R), survey of young person following consent before Your Choice starts (in some cases, some form of BAU work/support will have

		already been taking place). Measured as part of the baseline participant's survey, administered in week 1 after recruitment, before the start of Your Choice. *Internalising behaviours*: emotional difficulties and peer difficulties subscales of the SDQ measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Hyperactivity*: hyperactivity subscale of the SDQ measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Prosocial behaviours*: Strength and Difficulties prosocial behaviour subscale measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Prosocial identity*: Pro-social Identity Scale (PIDS) measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively. *Self-reported and practitioner-reported perceptions of young person's safety*: Young person and practitioner versions of "Checkpoint. A safety scale for young people", which is an instrument to measure young people's perceptions of safety developed by the research team. Measured as part of the baseline young person and practitioner surveys, administered at week 1 after recruitment.

Randomisation

Randomisation across the full study (internal pilot + efficacy)

Randomisation is **at the unit of teams of practitioners**. The rationale for this is that practitioners within the same team interact a lot with each other. Randomisation at the level of the practitioner would therefore create too much of a risk of cross-contamination.

Randomisation is **stratified at the LA level**. This model ensures that there will be trained and control teams in all participating LAs, a configuration that aligns well with the model of funding for delivery of the intervention.

Randomisation happened in three phases, that correspond to three batches of funding for programme delivery mentioned earlier: the Home Office phase (December 2021), the pilot phase (spring 2022) and the efficacy phase (summer 2023). The randomisation procedure was the same at all phases, to ensure continuity and in line with the internal nature of the pilot. Specifically, in an initial step LA put forward a list of so far untrained teams that they have available for delivering Your Choice. Among LAs that have two or more teams available for training, the evaluators randomly select teams for being trained and teams for being controls. The randomisation code is developed in STATA 16. The outcome of the randomisation is communicated to LAs by the evaluators, and the training of selected teams is implemented.

To check adherence to randomisation in the efficacy trial, we will require LAs to provide a list of practitioners in each of the teams involved in the study and to let us know if and when they have been trained ahead of start of delivery in the efficacy phase.

Practitioners will know whether they are part of a treated or control team and therefore cannot be blind to the experimental assignment. When recruiting young people, practitioners will know the young person's experimental condition. This means that the young person may not be blind to the assignment at the time of recruitment. Training of practitioners and SPOCs around recruitment will strongly emphasize the need to recruit young people into the study (as opposed to an experimental arm) and to not divulge their treatment condition.

Participants

Participating Local Authorities

Teams involved in the full evaluation (across internal pilot and efficacy) are from 29 LAs. 28 of these 29 LAs will be involved in the efficacy phase of the trial. Among the 4 LAs of London that are not participating in the efficacy phase of the trial:

- One (Enfield) recruited into a randomised team during the pilot but is not able to participate in the efficacy trial
- Three (Lambeth, Wandsworth, Camden) were involved in the pilot trial but either did not recruit anyone, did not have teams to randomise or did not adhere to the randomisation in the pilot trial so as to be included in the internal pilot. These three

LAs also were not able to put forward teams to be randomised in the efficacy trial. Therefore, they are not represented in either the internal pilot or the efficacy trial.

Although these LAs must be excluded from the experimental evaluation of Your Choice given that they would jeopardise the experimental design, data will be collected on young people in these LAs assigned to trained teams delivering Your Choice, and to control teams. That data will be used to expand the sample size for a parallel non-experimental evaluation providing further support to the experimental results, and to further characterise heterogeneity in delivering Your Choice in practice by taking into account the diversity of LA environments.

Participating teams of practitioners

LAs were asked to identify teams they would like to have participate in the efficacy trial as teams that meet the following two conditions:

a) frequently support young people aged 11-18 at medium to high risk of contextual harm

b) consent to train in and deliver Your Choice should they be randomly assigned to deliver the experimental condition.

Across the whole study, there are 130 teams involved in one of the three batches of the randomisation. As Table 2 details, these 130 teams include 7 teams that only participated in the pilot trial, 93 teams that will only participate in the efficacy trial, and 30 teams that will participate in both. In the latter group, the 16 teams that acted as controls in the pilot were pulled back into the randomisation pool for the efficacy trial. 8 of these teams were randomised to retain their control status. But 8 others were randomised to switch to treated status. From the point of the view of the randomisation (and power calculations), these teams amount to new teams in efficacy so while the number of unique teams is 130, the effective number of teams is 138 (since these 8 teams count as 8 controls during the pilot and as 8 treated teams during efficacy).

Table 2: Teams involved in the whole evaluation (including internal pilot)

Type of teams based on involvement in pilot and efficacy phases	Number of unique teams and breakdown between control and treated	Number of effective teams and breakdown between control and treated	Number of LAs represented

A. Teams that only recruited during the pilot trial	7 unique teams, including: • 3 control • 4 treated	15 effective teams including: • 11 control • 4 treated	5 LAs
B. Teams that are only expected to recruit during the efficacy trial	93 unique teams, including: • 47 control • 46 treated	101 effective teams including: • 47 control • 54 treated	25 LAs
C. Teams that both recruited during the pilot trial and are expected to recruit during the pilot trial	30 unique teams, including: • 8 control in pilot, control in efficacy • 8 control in pilot, treated in efficacy • 14 treated in pilot, treated in efficacy	24 unique teams, including: • 8 control in pilot, control in efficacy • 14 treated in pilot, treated in efficacy	18 LAs
Total across pilot and efficacy	**130 unique teams**	**138 effective teams**	**29 LAs**

Teams of practitioners involved in efficacy phase

We provide further details on the 123 teams (type B and type C above) involved in the efficacy phase of the trial and on the randomisation performed to allocate treatment to those teams whose condition could be randomised in this round. Table 3 below breaks down these 123 teams into six types, depending on their status in previous phases of delivery. It shows that 22 teams (of types 5 and 6) cannot be randomised in efficacy because they were randomised into treatment during previous batches of randomisation and so can only continue acting as a treated team in efficacy. Among the other 101 teams (of types 1, 2, 3, 4), randomisation cannot take place in LAs where there is only 1 team of this type. This happens in 6 LAs (Barnet,

Bexley, Havering, Hounslow, Lewisham, Newham). This means that we end up with 95 teams for randomisation in the efficacy stage of the study.

Table 3: Teams involved in the efficacy stage of the evaluation

Type of team	Number of teams
1. New teams (these teams were never included in the HO or pilot batches of randomisation, nor did they recruit any participant in previous phases)	69
2. Teams that were randomised into control during the Home Office phase	3
3. Teams that were randomised into control during the plot phase	28
4. Team that was non-randomly selected as control during the pilot phase	1
5. Teams that were randomised into treatment during the Home Office phase	4
6. Teams that were randomised into treatment during the pilot phase	18
Total	**123**

As for the pilot phase, the 28 LAs that are participating in the efficacy phase will be required to sign a Grant Agreement with the Violence Reduction Unit outlining the funding they will receive, as well as minimum requirements attached to their participation, including their willingness for their teams to partake in the randomisation, their readiness to share data about participants' background information and compliance with the study, their willingness to facilitate survey data collection, and the readiness to deliver Your Choice during the duration of the pilot.

Participating young people

Participants will be recruited among the young people allocated to be supported by participating teams who meet the following eligibility criteria:

a) age 11-18 years old (inclusive) at the time of recruitment;

b) at medium to high risk of contextual harm and referred to LA services with a view of mitigating such risk. This assessment will need to be quality assured by a MACE/pre-MACE panel or by the practitioner's team manager.

Recruitment of young people to the study will be done by the young person's lead practitioner. The practitioner provides information about the study using the information sheet and recruitment video. After discussing what participation would entail in one or several sessions, the practitioner will ask the young person to sign a consent form. For young people aged 11-15 years old, the practitioner will also involve the young person's parent in the recruitment discussion and require the young person's parent or guardian to sign a parental consent form.

Rationale for planned number of participants

The planned number of participants in the efficacy phase is 1,857. This number includes:

- 179 young people recruited in randomised teams during the pilot phase (among them, 83 young people have endline young people questionnaires completed at the time of writing).
- 1,678 young people expected to be recruited in randomised teams during the efficacy phase (Aug 2023 – December 2024).[5]

We believe the expected number of participants during the efficacy phase is realistic for the following reasons. First, these numbers have been discussed by the evaluation team in one-to-one conversations or email with each LA. Second, several LAs have purposely provided lower bounds to the numbers they think they can actually recruit in order to be realistic and successful in meeting these numbers. Third, the efficacy trial is planned to involve about five times as many randomised teams as the pilot trial did, and to last roughly 3 to 4 times as long as the pilot trial. We therefore think that it is reasonable for the number of participants in the efficacy trial to be at least 10 times higher than the pilot trial, or around at least 1,600 young people.

Settings of intervention delivery

[5] At the time of writing, 4 Local Authorities had not confirmed their expected numbers of participants and we therefore made conservative assumptions about their expected numbers of participants.

Sessions between young people and practitioners can take place in a variety of settings, from the young person's home, to a council building (e.g. a youth or community centre, school) to a public place (e.g. park).

Settings of data collection

Baseline and endline young people questionnaires will be collected during sessions between the young person and their lead practitioner (and potentially their parent or guardian). As mentioned above, these sessions can take place in a variety of settings. Young people will complete the online forms and questionnaires on the practitioners' work device which can be a phone, tablet or computer. Practitioners will fill out the baseline and endline practitioner questionnaires as well as session forms on their work devices.

Sample size calculations

Based on LAs' plans for the efficacy trial, the expected analytical sample will have two features that are important to consider for inference:

- *There will be a small number of units of randomisation (teams) within strata (LA)*: Among the 29 LAs participating in the whole evaluation (internal pilot + efficacy), the median number of teams per LA is 3 (the average 4.6), with all but 2 LAs having 10 or fewer teams involved in the trial (Brent has 12 teams and Barking and Dagenham has 21 teams).[6]
- *Units of randomisation (teams) have more than 10 observations on average*: Based on the conservative predictions of participating young people provided by Local Authorities so far, the average expected number of participants per team is 13.[7]

De Chaisemartin and Ramirez-Cueller (2022)[8] show that in stratified cluster RCT where the number of units of randomisation (teams) is small (10 or fewer) within strata (Local Authorities), the regular practice of clustering standard errors at the unit of randomisation can lead to downward bias in estimates of the variance of the treatment effect, resulting in over-rejecting the null hypothesis (of no effect). They advise that standard errors should be

[6] Even after removing the 6 teams that recruited in the pilot but did not collect any endline data, the median number of teams per LA is also 3 (the average if 4.5).

[7] After considering the actual attrition at endline in the pilot (54%) and the expected 30% attrition rate at endline in the efficacy, we would have 9.4 observations per team on average.

[8] De Chaisermartin, C. and Ramirez-Cueller, J. (2022). "At what level should one cluster standard errors and small-strata experiments?", National Bureau of Economic Research WP 27609, http://www.nber.org/papers/w27609

clustered at the strata level (LA) when RCTs have this type of configuration (and this result holds whether or not strata fixed effects are controlled for).[9]

We are not aware of any software allowing us to perform power calculations that take these issues into account. In order to account for the exact nature of our dataset and for these technical complexities, we therefore opted for calculating the Minimum Detectable Effect associated with 0.8 power via simulations programmed in STATA and adapting the procedure suggested in McConnell and Vera-Hernandez (2015)[10]. The simulation code used to perform these calculations is enclosed but we describe the steps used to perform these simulations below.

Each simulation is characterised by five parameters:

- Standard deviation of the outcome (σ)
- Intra-cluster correlation (ρ)
- Minimum Detectable Effect (β)
- Attrition rate (r)
- A threshold on the continuous outcome above which the binary indicator takes the value 1 and 0 otherwise (η)

Each simulation is based the following steps:

1. Randomise teams to be newly randomised to treated or control within each LA with 50-50 split.
2. In LAs that have an odd number of teams to randomise, we randomly select whether (n+1)/2 or (n-1)/2 teams would get treated (where n is the number of teams to randomise)
3. Generate a normally distributed random variable at the Local Authority level, θ_j, of mean 0 and variance $\rho\sigma^2$
4. Generate a normally distributed random variable at the young person level, ϵ_{ij}, of mean 0 and variance $(1-\rho)\sigma^2$
5. Implement the following data generation process for the outcome Y:

$$Y_{ij} = \beta T_{ij} + \theta_j + \epsilon_{ij}$$

where T_{ij} =1 if participant *i* is recruited in a treated team and 0 in a control team

[9] In cases where clusters are small (fewer than 10 observations per unit of randomisation on average), estimates of the standard errors should not adjust for degrees of freedom. This is not the case we expect even under the conservative projections in terms of participating young people.

[10] McConnell, B and Vera-Hernandez, M. (2015). Going beyond simple sample size calculations: a practitioner's guide. London: Institute for Fiscal Studies. Available at: https://ifs.org.uk/publications/going-beyond-simple-sample-size-calculations-practitioners-guide

6. Create a binary outcome D_{ij} such that it takes the value 1 if the score above is above a particular threshold η[11] and 0 otherwise[12]
7. Estimate OLS regressions of the chosen outcome (continuous Y_{ij} or binary D_{ij}) on the treatment dummy, clustering the standard errors at the LA level, with or without controlling for Local Authority fixed effects, on a randomly selected sample of observations of size (1-*r*) of the expected number of study participants in order to simulate a random attrition rate *r* to which we add the 83 observations from pilot that have endline data.
8. Repeat steps 1 through 6 1000 times and compute the power as the proportion of times the coefficient β is significant at the 95% level.

As part of this exercise, we wanted to explore the power implications of controlling for a lagged (baseline) outcome. Instead of making additional assumptions to simulate such baseline outcome, we instead performed the simulation above with smaller values of the variance of the outcome variable and of the ICC in order to mimic the effect of controlling for a lagged outcome on these parameters as estimated using data from the pilot period.

Table 4: Values picked for parameters in simulations and justification *(in italics)*

	Simulation 1 Not controlling for lagged outcome	**Simulation 2** Controlling for lagged outcome	**Simulation 3** Not controlling for lagged outcome	**Simulation 4** Controlling for lagged outcome
Attrition rate (r)	30%		40%	
	(pilot attrition rate is of 50%)			
Standard deviation of	1 *(standardised outcome so that β is*	0.5 *(estimated on relevant endline SDQ*	1 *(standardised outcome so that β*	0.5 *(estimated on relevant endline*

[11] This threshold is chosen such that the mean of D is equal to the probability of observing our primary outcome in the baseline pilot data.

[12] This step is to report power calculations for the binary outcome based on the continuous score, though the data generating process assumed here only enables us to simulate average impacts on the continuous score and hence may miss possibly stronger impacts the intervention may have at the top of the distribution

the outcome (σ)	*to be interpreted in units of SD)*	*score after controlling for relevant baseline SDQ and treatment variable)*	*is to be interpreted in units of SD)*	*SDQ score after controlling for relevant baseline SDQ and treatment variable)*
Intra-cluster correlation (ρ)	0.12 *(estimated on relevant baseline and endline SDQ score)*	0.05 *(estimated on relevant endline SDQ score after controlling for relevant baseline SDQ and treatment variable)*	0.12 *(estimated on relevant baseline and endline SDQ score)*	0.05 *(estimated on relevant endline SDQ score after controlling for relevant baseline SDQ and treatment variable)*
Minimum Detectable Effect (β)	A grid of values between 0.05 and 0.25, with 0.01 increment *(YEF wants to see trials powered with MDE of 0.2 and we included a grid large enough for the treatment effect on the continuous score to provide a large grid of implicit values for the treatment effect on the binary score)*			
Value of continuous score above which child is in the high to very high range (η)	70^{th} percentile of the continuous score, so that the probability of scoring in the high/very high range is 0.3 as in the control group in the pilot data *(This is the minimum value that young people in the high or very high range for either conduct problem or hyperactivity score on the standardised SDQ externalising score)*			

Primary population of interest

The primary population of interest will be young people age 11-18 referred to their Local Authorities for support and at medium to high risk of contextual harm. The Local Authorities' plans in terms of teams and expected numbers of participants in each team were used to simulate data on participants.

Table 5 provides information about key aspects of this data-set. The number of teams included in the power calculations is 132 (as opposed to 138) because 6 teams that recruited in the internal pilot do not have any endline questionnaire data on their participants. Since the primary outcome is from the endline questionnaire, we remove these 6 teams from the

data. We assume that all teams involved in the efficacy trial will recruit participants and collect some endline data.

The number of participants in the power calculations is 1761 (as opposed to 1,857 because it removes the 96 pilot participants that do not have endline data).

Table 5: Structure of the data used to perform power calculations

		PARAMETER
Average cluster size (if clustered)		Average number of participants per cluster (team): 13
Number of clusters (teams)	Intervention	70
	Control	62
	Total	132
Number of participants	Intervention	960
	Control	801
	Total	1761

Note that although randomisation during the efficacy phase will allocate teams to training and control on a 50-50 basis, legacy randomisation from the Home Office and pilot phases mean more teams are trained than control. It should be highlighted, however, that this imbalance across teams is expected to contribute to a more even distribution of young people across randomisation arms given that Your Choice demands more time being dedicated to each young person, so treated teams do not have the capacity to accept as many young people as control teams over a limited period.

The results of these simulations allow us to pin down the MDE under the below assumptions about power, significance level and the type of test we want to perform and under different assumptions about the attrition rate and about whether Local Authority fixed effects are controlled for in the regression.

Table 6: Minimum Detectable Effect on primary outcome (binary indicator for high to very high range of conduct problems) under the assumption we do not control for lagged outcome

Alpha[13]	0.05		
Power	0.8		
One-sided or two-sided?	Two-sided		
ICC	0.12		
Variance of continuous score	1		
Continuous score of externalising difficulties (standardised to have SD 1)	Attrition rate = 30%	No strata fixed effects	MDE = 0.17
		Strata fixed effects	MDE =0.16
	Attrition rate = 40%	No strata fixed effects	MDE =0.18
		Strata fixed effects	MDE =0.17
Binary score for scoring in high or very high range on conduct problems (probability of outcome in control group assumed to be 25%)	Attrition rate = 30%	No strata fixed effects	MDE = 24% of control mean
		Strata fixed effects	MDE = 23% of control mean
	Attrition rate = 40%	No strata fixed effects	MDE = 25% of control mean
		Strata fixed effects	MDE = 25% of control mean

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

Table 7: Minimum Detectable Effect under the assumption we do control for lagged outcome that reduces the variance of the underlying outcome to 0.6 and the ICC to 0.05

Alpha[14]	0.05		
Power	0.8		
One-sided or two-sided?	Two-sided		
ICC	0.05		
Variance of continuous score	0.6		
Continuous score of externalising difficulties (standardised to have SD 1)	Attrition rate = 30%	No strata fixed effects	MDE = 0.12
		Strata fixed effects	MDE = 0.12
	Attrition rate = 40%	No strata fixed effects	MDE = 0.13
		Strata fixed effects	MDE = 0.13
Binary score for scoring in high or very high range of the conduct problem scale	Attrition rate = 30%	No strata fixed effects	MDE = 24% of control mean
		Strata fixed effects	MDE = 24% of control mean
	Attrition rate = 40%	No strata fixed effects	MDE = 25 % of control mean
		Strata fixed effects	MDE = 25% of control mean

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

Discussion of MDEs in the context of the pilot trial and of the literature

The power calculations in Table 6 above indicate that, under a 30% attrition rate, we could detect with 80% power a minimum effect of 24% of baseline probability to be in the high to very high range of conduct problems.

This MDE is within the range of impacts suggested by the endline data well below the effects of CBT interventions on externalising behaviours, such as Attention Deficit Hyperactivity Disorder, Conduct Disorder and Oppositional Defiant Disorder, which are reported in the literature and which range around 60-70% (Gaffney, Farrington and White, 2021[OBJ]).

Using the pilot data on all teams (both randomised and non-randomised in order to maximise the sample size), we estimate that the impact of the intervention is a 34% reduction in the likelihood of a young person to be in the high to very high range of the conduct problem scale of the SDQ. (This coefficient is robust to the inclusion of the baseline outcome as control). Although this impact is not statistically significant, it is of higher magnitude than the MDE the whole study is powered to detect.

Overall, based both on the suggestive evidence from the pilot as well as from the literature on CBT interventions, we think the minimum effect the trial would be powered to detect is within the range of reasonable effects to expect from the Your Choice intervention.

Outcome measures

The baseline and outcome measures that will be used in this trial are described in Table 1. Baseline measures will be gathered shortly after recruitment in week 1, using instruments in the baseline surveys of young people and practitioners. Outcome measures will be assessed at endline, in week 20 after recruitment, using the similar instruments in the endline surveys of young people and practitioners.

Baseline measures

As described in Table 1 and below in relation to the outcomes, these include measures of internalising problems (emotional problems and peer problems), externalising problems (conduct problems and hyperactivity), prosocial behaviour, prosocial identify, social connectedness, self-reported and practitioner-reported perceptions of young person's safety. All are measured as part of the baseline survey of participants and practitioners, taken shortly after recruitment at week 1. The instruments are exactly the same as those used at in the endline surveys for primary and secondary outcomes, explained below.

Primary outcome

Our primary outcome is a measure of conduct problems based on scores from the Strengths and Difficulties Questionnaire (Goodman et al., 1997). To provide a rich picture of potential non-linearities in the impact and insights on whether young people in the high to very high zone of the scale benefit the most, we will measure conduct disorder by an indicator for the high or very high range on SDQ conduct problems subscale, where we will be using the fourfold categorisation provided by the SDQ developers in identifying high and very high ranges for these subscales. This outcome will be measured for all participants as part of the young person endline survey, taken at week 20 after recruitment.

The association between the SDQ conduct problems subscale and disruptive behaviour including violent and non-violent deliquency has been demonstrated in the existing literature (e.g Spaan et al 2023, and references therein). As explained earlier in this document (see discussion under Research Question 1), we opted for a categorical variable rather than the continuous index to better capture effects on the margin that is arguably the most relevant in assessing the success of Your Choice given the high pre-disposition for violent behaviour and rule breaking of individuals scoring in the abnormal range for this subscale. We expect the impacts of a violence reduction programme to vary by conduct difficulties given their association with violent behaviour. The alternative of considering the continuous conduct problems index for primary outcome would therefore run into the risk of diluting estimated effects of the impact of the programme and fail to reveal the impact for the sub-group that may benefit the most given the nature of the intervention.

Relative to actual measures of criminal and violent activity, the SDQ conduct problem scale has at least three advantages. First, information on this outcome is readily available at the end of the participation period while criminal activity takes time to (gradually) build up. Hence it is favoured in cases when there is urgency in learning about the impact of the intervention. Second, impacts may also be more visible on this measure than on criminal activity even in the medium run, given the low frequency of crime and the fact that conduct problems are likely at the forefront of the causal chain of effects of the intervention. In a setting where the size of the experiment may compromise what can be identified, it is crucial to identify a primary outcome that promises that impacts can be detected. And third, it is a more adequate measure to capture effects both on violent behaviour and on vulnerability to be a victim of violence. Vulnerable youth is also a target group of Your Choice and one for which crime data may not provide insights on the impact of the programme.

Secondary outcomes

Secondary measures are meant to capture effects on dimensions that relate to reductions in violent behaviour, either because they are simultaneously affected by the intervention and support/reinforce effects on violent behaviour, or because they may reveal some of the

mechanisms through which violent behaviour is affected in the short and longer terms. We add a measure of offending behaviour to the list of secondary outcomes, which directly informs on the main goal of the intervention of reducing violence.

Criminal activity: measured by an indicator for a recorded arrest during the period of 16 months following recruitment. This time lag provides some scope for criminal activity to happen and be registered in policy records, which explains why this one outcome is measure later than all others. The PNC measure of criminal activity is an objective measure of offence activity and provides direct evidence on the key goal of the programme of reducing violence among young people most at risk of violent behaviour. Two features of these data explain that it is considered as a secondary outcome. First, the necessary delays in obtaining data on criminal activity, which at the earliest will be available in full only 2 years after the end of the trial, significantly extending the time required to produce key findings and learnings. And second, the low frequency of detectable criminal activity significantly increases the risk that the study would be under-powered to identify any effects given current projections of the size of the trial.

Data on recorded offences will be obtained from the Policy National Computer database, and linked to experimental data through the use of participant's name, LA and date of birth.

Self-reported and practitioner-reported perceptions of young person's safety: measures taken from the young person and practitioner versions of "Checkpoint: A safety scale for young people", which is part of the endline young person's survey taken in week 20 after recruitment and which was co-developed by the project and evaluation teams during the pilot phase of this project. Subjective measures of propensity to commit violent acts or of being vulnerable to being a victim of violence, provides an alternative view on the impact of the intervention on violent behaviour. Since it can signal changes in attitudes towards violent behaviour, these measures may also be especially informative about long-lasting effects of the intervention.

Wellbeing: measured by the Short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS), which is part of the endline young person's survey taken in week 20 after recruitment (Ng Fat et al. 2017). Feeds directly into the logic model as one of the expected outcomes from the intervention.

Emotional self-regulation: measured by the Self-regulation subscale from the Trait Emotional Intelligence Questionnaire – Adolescent Short Form (TEIQUE-ASF), which is part of the endline young person's survey taken in week 20 after recruitment. Captures aspects of emotional difficulties that the logic model predicts will be affected by the intervention.

Social connectedness: measured by the Social Connectedness Scale – Revised (SCS-R), which is part of the endline young person's survey taken in week 20 after recruitment. This scale will

help measure whether the intervention increases the young person's resilience through strengthened sources of social support (at home, at school if relevant, and elsewhere).

Internalising behaviours: emotional difficulties and peer difficulties subscales of the SDQ measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively.

Hyperactivity: hyperactivity subscale of the SDQ measured at baseline and endline as part of the participant's surveys administered following recruitment and at week 20 after recruitment respectively.

Prosocial behaviour: measured by the prosocial scale of the Strengths and Difficulties Questionnaire, which is part of the endline young person's survey taken in week 20 after recruitment. This measure will provide insight into whether the impact of the intervention may be partly driven by changes in prosocial behaviour, as hypothesized in the logic model.

Prosocial identity: measured by the Pro-social Identity Scale (PIDS), which is part of the endline young person's survey taken in week 20 after recruitment. This measure will provide insight into whether the impact of the intervention may be partly driven by changes in prosocial identity, as hypothesized in the logic model.

Data

For completeness and clarity it is worth briefly overviewing the data that will be available and the timing of data collection. More details can be found in the pilot report, with a copy of all questionnaires and of the session form available in Annex E of that report.

LA Workbook*:* Spreadsheet to be filled out by the SPOC/Data Lead of the Local Authority to provide background information about the young person, as well as a log of training dates for practitioners involved in the study and a log of clinical supervision sessions. The background information about the young person includes: date of birth, gender, ethnicity, SEN status, reason for referral into the service and any other services the young person is involved with. The workbook also includes a tab to list some basic demographic information (ethnicity, gender, age) about the young people who were approached to enter the study but who did not consent to and provides an option to explain the reason (if known) for non-consent.

Baseline young person questionnaire**:** The baseline young people survey is administered shortly after recruitment, in week 1, during a session with the assigned practitioner. It includes the Strengths and Difficulties Questionnaire (SDQ), a scale developed by the evaluation and project team about the young person's safety (young person Checkpoint), the Social Connectedness Scale – Revised (SCS-R), the Pro-Social IDentity Scale (PIDS) and a question about the young person's main activity at the time of the survey.

Baseline practitioner questionnaire: The baseline practitioner survey is completed shortly after recruitment, in week 1. It asks the young person's practitioner to report their risk rating of the young person, as well as their perception of the young person's safety using the practitioner version of Checkpoint.

Endline young person questionnaire: The endline young people survey is administered at the end of Your Choice period, in week 20 after recruitment, during a session with the assigned practitioner. It includes the same content as the baseline young person questionnaire, plus the following scales: The Short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS), the Trait Emotional Intelligence Questionnaire – Adolescent Short Form (TEIQUE-ASF) – Self regulation subscale, and the Pro-Social IDentity Scale (PIDS). The latter were excluded from the baseline survey to keep it as short as possible while focusing on the most relevant measures, and also because these scales are most interesting as outcomes rather than as conditioning variables given the rich data we already collect at baseline. The endline questionnaire also asks the young person some questions about their experience working with their practitioner: it asks whether they would recommend it to someone else, it provides a free text box for any additional comments, and it asks a series of 7 questions about the practitioners' approach and practices to working with the young person. These approaches/practices were selected to be those most characteristic of the practices embedded in the Your Choice training.

Endline practitioner questionnaire: The endline practitioner survey is completed at the end of Your Choice period, in week 20 after recruitment. It includes the practitioner version of Checkpoint, as well as a series of 7 questions about the practitioners' approach and practices to working with the young person. These approaches/practices are the same as the ones asked to the young person in the endline questionnaire and were selected to be those most characteristic of the practices embedded in the Your Choice training.

Session forms: Practitioners are requested to fill out a session form every time they conduct a session with the young person or every time a session is cancelled. The session form asks a few details about the location of the session, the young person's level of engagement, whether the session was related to the Your Choice programme, and depending on the latter offers one or two dropdown menus to describe the content of the session. The practitioner is then offered a free text box to include any further comments on the session. If the session did not take place, the session form asks whether the reason for cancellation is known and offers a free text box to include any other details about the session that did not take place.

Compliance

In the Your Choice trial, compliance is defined at three levels: LA, team and young person.

At the LA level, where randomisation is to be implemented within LA across teams:

- That the assignment of young people to trained and control teams is not selective on the young person's characteristics or aspects of their case. As mentioned below, we will not be able to fully rule out selective assignment of young people to teams, but we will obtain quantitative and qualitative evidence to support any claims of non-selective assignment – in particular, we expect to see balanced distribution of observed characteristics;
- That trained and control teams undertake similar efforts to recruit young people to participate in the trial – hence we would expect to see similar consent rates into treated and control teams.

At the team level, the following criteria define compliance:

- That in teams randomised to deliver Your Choice, all practitioners receive training (and when a new practitioner enters the team, they receive training to deliver Your Choice);
- That in teams randomised to the control group, no practitioner receives training;
- That trained practitioners intend to meet three times a week with the young people enrolled in Your Choice for 12 weeks (over the course of 12-18 weeks)
- That trained practitioners meet at least monthly with their clinical supervisor to discuss progress and the cases assigned to them;
- That control teams do not deliver elements of Your Choice in their interactions with young people assigned to them (other than what they would do in regular BAU;

At the young person level, the following criteria define compliance with the trial protocols:

- That the young person is not transferred across trained and control teams while participating in the programme;
- The young person undertakes the intended programme of support with the Local Authority and does not drop out of this programme prematurely
- That young people in the Your Choice meets 3 times a week with their practitioner for 12 weeks (over the course of 12 to 18 weeks)
- That the young person completes the planned surveys.

Information on compliance: Compliance with the randomisation of teams will be directly observed in the administrative data provided by the Local Authorities about the names and teams of practitioners involved in the study and their date of training if relevant. Information collected in session forms will reveal adherence to the delivery of Your Choice or BAU across teams, as well as violations to the rule that young people should not be re-allocated to other teams during the participation period. Qualitative data will be collected from a subsample of practitioners and clinical leads, to inform on the frequency and content of meetings between practitioners and their clinical supervisors. The baseline questionnaire will provide information to test for the presence of systematic differences between young people assigned to treated and control teams, either due to non-random allocation of young people

to teams or to differential selectivity in young people across treatment arms who agree to share their data (we will not be able to distinguish between the two sources of bias).

Analysis

Our primary analysis uses only randomised teams during the Home Office, pilot and efficacy phases. It will involve various steps.

Step 1. Thorough check that the group of young people assigned to Your Choice appear to come from the same population as those assigned to the control group along many observed variables describing demographic characteristics (e.g. gender, age, ethnicity, etc), main activity at baseline (education or training, employment, other), whether looked after, special education needs, primary needs for Local Authority support (e.g. abuse, disability, family acute stress, etc), various measures observed in the baseline youth survey including SDQ scores and self-reported risk assessment, practitioner risk assessment from baseline practitioner questionnaire. We will also check that the distribution of services to which young people have been assigned look balanced across experimental arms. All subsequent analysis addressing the key research questions set out earlier in this document must rely on a good understanding of the comparability between the treated and control groups.

Step 2. Thorough check of the delivery of Your Choice in contrast with BAU, revealing the extent to which the programme effectively changes interactions between the youth workers and young people and characterising the intensity of treatment for different groups. Specifically, we will check that young people assigned to Your Choice received an intensive schedule of meetings with content that is specific to the programme (hence matching the Your Choice protocol), while young people assigned to the control teams did not. Again, all subsequent analysis and interpretation of findings must rely on a good understanding of what the programme delivers in comparison with BAU.

Step 3. Main analysis, which will be a linear probability model to measure the impact of Your Choice on an indicator for high to very high conduct problems (primary outcome). The regression will also control for the baseline outcome and any controls for which we find statistically significance imbalances in step 1, as well as Local Authority fixed effects. Standard errors will be clustered at the Local Authority level, following Chaisemartin and Ramirez-Cuellar (2022). Effects will be shown in percentage points and as a percent change relative to the baseline average frequency of young people in the high to very high range for this population. We will show the 95% confidence interval for the estimated effect. This analysis will directly address research question 1.

Step 4. Robustness analysis of estimated effects on primary outcome. This will involve estimating alternative versions of the primary regression model using a post-double selection

LASSO to identify variables that are related with both the treatment status and the primary outcome. We will consider relevant contextual, demographic, personal and environmental information as described in step 1. This analysis will provide further support for the findings in the main analysis (step 3).

Step 5. Steps 3 and 4 will be repeated for the secondary outcomes measuring whether the young person got arrested at any point during the 16 months following recruitment into Your Choice (PNC data), and risk assessments by the young person and practitioner at endline (survey completed at week 20 after recruitment). Effects on arrest probabilities will be shown in percentage points and as a percent change relative to frequency of arrests among the control group over the same period. Effects on risk levels will be shown in standard deviation units. Standard errors will be clustered at the local authority level and we will show the 95% confidence interval for the estimated effect. This analysis will address research question 2.

Step 6. Investigation of potential mechanisms through which Your Choice affects externalising behaviour, arrests or risk assessment. We will regress each of the other secondary outcomes measured at endline (week 20 after recruitment) on a treatment indicator, the baseline value of that variable where available, other observed characteristics where imbalances at baseline were found (steps 1 and 2), and local Authority fixed effects. This will provide suggestive evidence on potential mechanisms underlying identified responses in externalising difficulties and criminal behaviour and risk, as detailed in the Logic Model. Effects will be shown in standard deviation units. Standard errors will be clustered at the local authority level and we will show the 95% confidence interval for the estimated effect. This analysis addresses research question 3. We will not run a fully fleshed mediation analysis because we do not have exogenous variation to independently vary the mediators.

Step 8. Heterogeneity analysis of the impact of Your Choice on the primary outcome (conduct problems) and criminal activity (possibility of arrest). This will be carried out by adding interaction terms between the treatment indicator and variables describing key demographic and contextual factors. Specifically, we will interact the treatment indicator with age, gender, ethnicity, special education needs and practitioner risk assessments at baseline. This analysis will address research question 4.

[15]

Longitudinal follow-ups

No follow-up points have been agreed at this stage.

[16] Ritchie, J. and L. Spencer, Qualitative Data Analysis for Applied Policy Research, in Analyzing Qualitative Data, A. Bryman and B. Burgess, Editors. 1994, Routledge: London.

Implementation and process evaluation

Research questions

The research questions of the implementation and process evaluation (IPE) are:

1. To what extent is Your Choice delivered as intended?
2. What are the barriers and facilitators to delivery?
3. What are young people's and practitioners' views and experiences of Your Choice (acceptability, impacts, and mechanisms)?
4. What are practitioners' views on the sustainability of Your Choice?

Research methods

To address the above research questions, we will conduct a mixed-methods IPE. We will use a triangulation design convergence model, where quantitative and qualitative data will be collected and analysed separately and then interpreted together to provide complementary perspectives on the same topic.

The quantitative data will involve analysis of session survey forms (tested during the pilot) completed by a practitioner after each session with a young person in both Your Choice and BAU arms.

The qualitative data will focus on the Your Choice arm and will involve recruiting young people (*n* = 16-20), Your Choice trainers (*n* = 8-10), youth practitioners (*n* = 16-20), and other professionals (e.g., multi-agency risk or exploitation panels, service managers) (*n* = 9-12). The sample sizes were determined based on the research team's experience of conducting similar previous studies to achieve diversity in views and perspectives whilst being achievable within the resources of the project. Interviews/focus groups with youth practitioners will be conducted during the mid-stages of the trial. Interviews with young people will be conducted twice, on completion of the programme and approximately six months later, to examine sustainability of short-term changes and any longer-term changes. Topics that will be examined in the interviews/focus groups build on those tested in the pilot. Topics will illustrative example questions include:

- Views and experience of Your Choice
- To what extent does the programme fit with and add to the landscape of existing practice?
- How acceptable is Your Choice?
- What could be done differently to make the sessions better for young people?

How the programme is implemented

- To what extent does this fit, or not fit, the theory of change and intervention description?
- How, if at all, could the programme be tailored to meet the context and population needs?

Barriers and facilitators to implementation

- What helps and hinders recruiting young people to the programme and then engaging them?
- Which components of the intervention are more, and less, readily delivered?

Impact of Your Choice on young people

- Is there evidence to support anticipated outcomes in the logic model?
- Are there any unintended consequences (e.g., due to continued contact with the Youth Justice System)

Sustainability

- What are professionals' views on how to continue delivery beyond the end of the project?

Analysis

Table 8: IPE methods overview (adapt as necessary)

Research methods	Data collection methods	Participants/ data sources (type, number)	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
IPE	Survey	Session forms completed by the practitioner after each session of Your Choice or BAU	Descriptive statistics and inferential statistics (e.g., t-tests, chi-square, regression) of the session forms (described in the impact	IPE research questions 1 and 2	Intervention description, logic model

			evaluation section; p.39)		
IPE	Interview	16-20 young people (Your Choice arm)	Framework approach and thematic analysis	IPE research questions 1, 3, and 4	Intervention description, logic model
IPE	Interviews/ focus group	Your Choice trainers (n = 8-10), youth practitioners (n = 16-20), and other professionals (e.g., multi-agency risk or exploitation panels, service managers) (n = 9-12)	Framework approach and thematic analysis	IPE research questions 1, 3, 4, and 5	Intervention description, logic model

Note. IPE = implementation and process evaluation.

The quantitative data will be analysed to answer IPE research questions 1 and 2:

1. To what extent is Your Choice delivered as intended?

Using descriptive statistics, we will examine the number and type of sessions and duration of support in the Your Choice arm, which will be compared with the intervention description.

2. To what extent is delivery of Your Choice different, and similar, to BAU?

We will run similar descriptive statistics for the BAU arm. We will then conduct inferential statistics (e.g., t-test, chi-square tests) to examine differences in the number and type of sessions and duration of support between Your Choice and BAU using the process data and workbook data.

Qualitative data (i.e., transcripts, free-text responses) will be analysed using the NVivo qualitative data analysis software. We will use the framework analysis approach[16] to manage

[16] Ritchie, J. and L. Spencer, Qualitative Data Analysis for Applied Policy Research, in Analyzing Qualitative Data, A. Bryman and B. Burgess, Editors. 1994, Routledge: London.

the data, categorising transcripts according to which component of the logic model they address. We will then use thematic analysis[17] to analyse the data organised in the framework to explore themes across participants' experiences and perspectives. At least two members of staff (including the peer researcher) will be involved and there will be regular coding review meetings throughout the stages of the analysis. Themes will be identified using inductive and deductive approaches, drawing on the logic model but also any new themes or ideas that are present in participant responses. Such approaches are commonly used in applied policy evaluations. Different reliability processes are available for qualitative data than quantitative data, and the research team will adhere to quality standards for establishing the trustworthiness of the data (i.e., credibility, transferability, dependability, and confirmability).[18]

The qualitative data will be used to address IPE research questions 1, 3, 4, and 5:

1. To what extent is Your Choice delivered as intended?
3. What are the barriers and facilitators to delivery?
4. What are young people's and practitioners' views and experiences of Your Choice (acceptability, impacts, and mechanisms)?
5. What are practitioners' views on the sustainability of Your Choice?

Young people's and practitioners' descriptions of Your Choice will be integrated with the quantitative data and intervention descriptions. This will enable us to, for example, gain some understanding of why there might be differences in the quantitative data between what was delivered and what was expected based on the intervention description. Understanding of barriers and facilitators to delivery will also enable us to further understand what was delivered and why it was delivered that way.

Young people's and practitioners' views and experiences of Your Choice in relation to acceptability, impacts, and mechanisms of change will be interpreted with, and complement, data from the overall quantitative efficacy trial on outcomes and mechanisms. Practitioners' views on the sustainability of Your Choice, and barriers and facilitators to sustainability, will provide evidence to inform potential plans for the continued delivery and scale up of Your Choice.

One of the outputs of the qualitative component will be vignettes, co-produced with the Youth Advisory Panel, based on participants' collective experiences of the programme. The

[17] Braun, V. and V. Clarke, Using thematic analysis in psychology. Qualitative Research in Psychology, 2006. 3(2): p. 77-101.

[18] Yardley, L., Dilemmas in qualitative health research. Psychology & Health, 2000. 15(2): p. 215-228.

aim of these vignettes is to articulate young people's journey through the intervention. These will be used to help communication about the key aspects and findings of the project in accessible ways for public-facing dissemination (e.g., public summaries). A purpose of the vignettes is to help anchor the evaluation findings to the views and experiences of the young people.

In the final report, we will integrate findings from the quantitative and qualitative components of the full trial[19], which involves collecting complementary quantitative and qualitative data on the topic to answer the research questions. We will use meta-matrices to integrate findings,[20] which involve summarising findings from different data sources according to the research question they address.

Cost data reporting and collecting

The organisations delivering the intervention are Local Authorities of London and will be asked to report on the cost of implementing Your Choice during the efficacy trial. We will aim to collect data from all the 28 Local Authorities in London. We will aggregate the reported costs across all of them to produce an average cost of implementation. We will report the fraction of overall cost coming from set-up versus recurring costs, as well as the fraction of overall cost coming from different types of costs (Staff, material and equipment).

To collect such data, we will use a template, which was developed following YEF guidance and in conjunction with the project team and which was piloted during the pilot trial with a number of Local Authorities.

The Local Authorities will be asked to report cost for a `typical cohort' defined as 3 young people supported by a team of 6 practitioners (5 practitioners and one team leader) through the duration of the programme (the equivalent to 3 sessions per week for 12 weeks), including training costs. Full compliance among these three young people will be assumed.

Diversity, equity and inclusion

Several steps have been taken to ensure the evaluation is accessible to all, welcoming and inclusive. This is important to ensure that the evaluation is representative of the target population, that participation is maximised, and that those approached feel that the

[19] Creswell, J.W., et al., Advanced mixed methods research designs. , in Handbook of mixed methods in social and behavioral research, A. Tashakkori and C. Teddlie, Editors. 2003, Sage: Thousand Oaks, CA.

[20] Wendler, M.C., Triangulation using a meta-matrix. J Adv Nurs, 2001. 35(4): p. 521-5.

programme and the evaluation efforts are inclusive and fair. Concretely, steps in this direction include:

- The development of recruitment material as well as questionnaires received the VRU Young People Advisory Group, a research Young People Advisory Group (gathered by the evaluation team), and a Speech and Language Therapy specialist from one of the Local Authorities involved in the trial. The YPAGs include individuals with lived experiences that are similar to those of participants in the study.
- In order to enhance the accessibility of information sheet and privacy notice, we have developed two 2-min animation videos to support the recruitment process and that summarises the key information in these two documents. These videos are intended to be shown by the practitioner to the young person at recruitment.
- In order to enhance the accessibility of questionnaires, we have designed the questionnaires such that bubbles with definitions of difficult words (in validated scales, where we cannot change the text) pop up when hovering over these words. Moreover, the questionnaires will include the option to have questions read out loud.

Ethics and registration

We will submit a high-risk ethics application to the University College London Research Ethics Committee (UCL REC) on 5th June. Ahead of this date, we will submit a Data Protection Impact Assessment to IFS and AFC Data Protection Officers to obtain data registration.

Data protection

The evaluation team will treat data protection during the efficacy trial with utmost consideration. It already developed comprehensive data information governance documentation to clarify data flows and outline the legal framework allowing for the collection, sharing, storing and processing of the data gathered to meet the research objectives of the study.

The evaluation team will develop a privacy notice for young people and their parents, and a privacy notice for practitioners to collect information about them through surveys and the LA workbooks. This privacy notice will strongly build on the privacy notice developed for the pilot trial, during which the project and evaluation teams worked with the Information Governance for London (IGfL) group to develop a data sharing agreement between the LA and the evaluator to allow LAs to share administrative data on participants and practitioners through the LA workbooks. The information sheet for young people (and their parents) as well as the recruitment animation video summarised key elements of the privacy notice.

The information collected during the trial will be stored by the IFS and/or by the AFC for the purposes of this project. These data, for most the most part, will not be anonymous, but both organisations have strong measures in place to ensure that only the research team can see the information from individual participants. Data is stored on the network of IFS in a secure folder with access restricted to named researchers. The IFS information security management system is ISO27001 compliant and the IFS has an Information Classification and Handling Policy which sets out a comprehensive set of guidelines for handling all types of data and information (including highly confidential information). AFC have similar Information Governance policies, and all information will be held on the secure AFC servers, with only approved researchers having access. All project team members will follow strict procedures in this policy and adhere to the IFS/AFC Information Security Policy when using or collecting data. All project team members have received appropriate GDPR training.

The only time someone other than a researcher in the evaluation team would see identifying information about participants alongside the information provided by the participants through questionnaires or interviews is if the participant answer to some survey questions, or interview responses, indicated that the participant or someone else would at risk of harm. In the efficacy trial, we will follow the safeguarding procedures that were developed during the pilot trial and that were approved by UCL REC and followed throughout the pilot trial.

The privacy notice developed during the pilot describes the complex data sharing processes associated with a) the linkage of primary data collected during the study to administrative records from the Department for Education and from the Ministry of Justice, and b) the archiving of data in the YEF data archive. In collaboration with the project team, the evaluation team strove to make these explanations as accessible as possible through the use of various diagrams. We plan to use a privacy notice that is very similar to the one used in the pilot trial, with the exception that the LIIA data lead (and an additional person to replace them during periods of leave) will be added as a processor of the data collected in order to support the collection of administrative data from LAs.

Data subjects' rights Data subjects have the right to ask for access to the personal information the evaluation hold about them, ask them to correct any personal information which is incorrect, and to erase the personal information when there is no good reason for continuing to hold it, although there are certain time limits lo for requesting deletion linked to the YEF data archive. The privacy notice provides the evaluation team's contact information for participants to get in touch.

Purposes for data processing The information sheet and privacy notice also clearly specify the purposes for data processing, as well as the parties with access to data and the reasons why in great details. It also states that the results of the research will be made publicly available through reports and presentations posted on the YEF, IFS and/or AFC websites.

Data retention The privacy notice also specifies that the data stored by IFS which includes the information from questionnaires as well as crime record from MoJ would be stored for a minimum of 10 years in order to allow the evaluation team to look at the longer-term effects. After 10 years, the IFS would carry out a review to see if there is still useful work that can be done using your data, committing that they would delete it at any point that they no longer need the data for this research project.

Data processing roles As specified in the privacy notice, the IFS and AFC are joint data controllers at the start of the study. When the study is finished in the second half of 2024, the data would be handed over to the Youth Endowment Fund (YEF) for archiving purposes, making YEF another data controller of the data. At this stage, the AFC will no longer become a data controller. A joint data controlling agreement is in place between the two institutions.

In the efficacy trial, a data processing agreement between IFS and the data lead at LIIA will be established in order to allow this person (and someone in their team to act on their behalf during periods of leave) to support Local Authorities with their completion of such data.

A data procession agreement between IFS and the data lead or Single Point of Contact (SPOC) in each Local Authority will also be established in order to allow these individuals to support practitioners with their duties regarding completion of consent forms and questionnaires.

Lawful basis The lawful basis for processing and storing the information on young people and practitioners during the study is the evaluation team's Legitimate Interest to research into the best way to support young people. By maintaining the YEF archive and allowing approved researchers to access the information in the archive, the YEF is performing a task in the public interest and this gives the YEF a lawful basis to use personal information. The lawful basis condition for processing "Special Category Data" (ethnicity, wellbeing/mental health data) is Article 9 (2) (j) Archiving, research and statistics. Processing crime record data in the project is done under Article 10 and the condition for processing that applies in this study is Research.

Stakeholders and interests

Our project is a partnership between the Institute for Fiscal Studies (IFS) and Anna Freud Centre (AFC). Professor Rasul, as PI, will have overall responsibility for the project. The IFS team lead the quantitative evaluation. The AFC team lead the qualitative process evaluation.

Rasul will lead engagement with YEF and the integration of quantitative and qualitative work streams. Cattan will lead in liaising with the delivery partner. All members will be engaged in the design of the evaluation and survey instruments. Costa-Dias will lead on methodological aspects and trial design. Cattan will lead on data collection and administrative data acquisition and supervise Nodee, who will lead on data cleaning, coordination of endline data collection, and compliance with safeguarding and voucher procedures.

Edbrooke-Childs will act as Implementation and Process Evaluation Lead. He will co-lead engagement with YEF alongside Rasul. Edbrooke-Childs and Stapley will lead on the methodological design. Jacob will work closely with the team on operational oversight, planning, and risk/issue log monitoring. Jacob and Labno will supervise the Peer Researcher (Orchard), with specialist input from Stapley and Edbrooke-Childs throughout. Labno will work on project management (support by Jacob); data collection, analysis, and reporting; and the Young People's Advisory Group and Peer Researcher involvement. All team members will analyse the data, with Edbrooke-Childs, Stapley, and Jacob leading the reporting and dissemination. Deighton will provide ongoing critical appraisal with a view of the overall process evaluation.

The evaluation team will continue to actively engage with the London Violence Reduction Unit and the London Innovation and Improvement Alliance in the co-design of the evaluation design, in communications to Local Authorities about the pilot trial, and in the design of a survey instrument to measure perceptions of young people's safety.

The evaluation team will continue to consult a research Young People Advisory Group.

Risks

We list the risks to the evaluation, with their likelihood of occurring and likely magnitude of impact, as well as how they might be addressed. We break down these risks into three areas: design, delivery and implementation and data collection Table

Table 9: Register of risks to trial design

Description of risk	Likelihood of risk	Impact of risk on trial success	Mitigation(s)
LAs do not provide enough teams to sufficiently power the efficacy trial	Low	High	Ask LAs to confirm list of teams to randomise and to have list signed-off by LA DCS before randomisation is performed. Trial will not start if list of signed-off teams does not meet required numbers
LAs randomise the wrong teams	Medium	Medium	Co-produce with project team clear communication strategy with LAs for the whole project, including: - Clear communication about teams to randomise for efficacy trial - 1to1s with LAs after randomisation plans confirmed to ensure full understanding of LAs - Regular check-ins by project and

			evaluation teams with LA SPOC about adherence to randomisation
LAs allocate young people into teams not following the designed assignment rule (team capacity and historical assignment for returning young people)	Low	Medium	Collect historical administrative anonymous data on allocation of young people to teams within services involved in study in a set of LAs, in order to check whether characteristics of young people
			Collect high quality baseline information from LA workbooks and baseline surveys
			Change the person making the allocation of young people to teams more distant from the study OR make the person recruiting the young people blind to the team the young person is allocated to
LAs recruit numbers of participants lower than expected	Medium	High	LAs to provide real-time information about their recruitment numbers
			VRU to build regular checks on recruitment numbers, with funding implications when targets are not met
			Provide particular support to LAs with higher (riskier) ratings but potentially large number of participants to ensure adherence to randomisation and data protocols
			Consider extending the period of recruitment beyond December 2024
Recruitment larger in treated teams than in control teams	Medium	High	Build better engagement of control teams into the evaluation
Other programmes running at the same time (e.g. Turnaround), reducing LA capacity to deliver Your Choice as part of YEF evaluation of Your Choice	Medium	High	VRU to build regular checks on recruitment numbers, with funding implications when targets are not met
Evaluation team stretched due to multiple LA troubleshotting and high admin requirements, alongside continuous data quality checks and efforts to increase endline completion rates	High	High	YEF to provide additional funding for a project manager hired and supervised by IFS evaluation team and to work across evaluation and project team
Sample size too low for trial to be	Medium	High	LAs to be provided real-time information about their recruitment numbers

powered to detect impacts on binary outcomes, such as offending, measured in PNC			VRU to build regular checks on recruitment numbers, with funding implications when targets are not met
			Provide particular support to LAs with higher (riskier) ratings but potentially large number of participants to ensure adherence to randomisation and data protocols
			Consider extending the period of recruitment beyond December 2024

Table 10: Register of risks to delivery and implementation

Description of risk	Likelihood of risk	Impact of risk on trial success	Mitigation(s)
Low buy-in of LAs staff into Your Choice	Medium	Medium	Appropriate communication and engagement strategy with DSCs and operational leads/service managers to ensure strong leadership buy-in Targeted communication strategy for practitioners about the programme and evaluation (e.g. webinars about the study so far, webinars and drop-ins with guidance for practitioners), with special attention to improving the buy-in of control practitioners
Capacity of practitioners and clinical leads constrained due to high intensity of Your Choice	Medium	Medium	Increase motivation and buy-in of practitioners and clinical leads by sharing more information about the study and improving communication strategy to make them feel more part of the study
High practitioner turnover, leading to need of retraining and delivery below expectations	High	Medium	Training of trainer enables training of new stuff to be done relatively quickly LAs encouraged to attend network groups to find cross-LA solutions Increase regular communication between evaluation team and LAs in order to support LAs find solutions quickly to capacity issues creating a risk to the trial

Table 11: Register of risks to data collection

Description of risk	Likelihood of risk	Impact of risk on trial success	Mitigation(s)

Workbooks not fully completed for young people in study	Low	Medium	Make DPA between LA and evaluator a requirement of participation in the efficacy trial so that evaluator can communicate the list of young people recruited into the study in real-time Have a DPA between LIIA data lead and IFS so that LIIA data lead can access evaluation data and provide more targeted support to LAs in completing workbooks Improve process to share workbooks through London Data Store from July 2023
Key variables (DOB, UPN) for admin data linked not provided by LAs and associated risk to data archiving	High	High	Have a DPA between LIIA data lead and IFS so that LIIA data lead can access evaluation data and provide more targeted support to LAs in completing workbooks
Delays in accessing administrative data for participants	Medium	High	Start PNC and NPD data applications at the start of efficacy trial to link data on pilot participants and set up an arrangement for linking batches of new efficacy participants every 6 months
High rate of non-consent among eligible young people (and their parents)	Medium	High	Focus groups with study participants to understand barriers to recruitment and improve recruitment materials Learning group of practitioners to share successful practices and develop evaluation toolkit for other practitioners Review of recruitment materials with young people advisory groups and peer researchers
Survey data of poor quality	Low	High	Ensure adequate team of peer researchers is appropriately funded to support young people with data collection
Low endline data completion rate	High	Medium	Explore possibility of increasing incentives for completing endline surveys with YEF and/or VRU additional funding Consult with VRU and Evaluation young people advisory groups as well as LA reference group, to improve endline data collection success Create mobile-optimised interface for

			practitioners and SPOCs, through which they will be able to: - Visualise, for each study participant in their LA, completed and missing data - Book endline data sessions with peer researchers Make primary outcome of trial based on PNC
Practitioner surveys not completed	Medium	Medium	Create mobile-optimised interface for practitioners and SPOCs, through which automatic reminders for completing missing surveys will be sent to practitioners and SPOCs
Session forms not duly completed	High	Medium	Create mobile-optimised interface to collect baseline survey data and session forms, which will: - Provide a real-time tracker of completed and missing data - Reduce time to complete each form by avoiding the need to re-enter information about the young person and practitioner on each form - Enable practitioners and SPOCs to visualise a log of session forms completed and to extract session form data for their own records - Allow SPOCs or data leads to complete information not completed by practitioner based on case notes

Timeline

Dates	Activity	Staff responsible/ leading
May 31 2023	Randomisation of teams into intervention	IFS
5th June 2023	Submission of ethics application to UCL REC	IFS, AFC
June-July 2023	Local Authorities train teams selected to deliver Your Choice as part of efficacy and refresh training of previously	Las

	trained teams partaking in the efficacy phase		
May-June 2023		Engagement of LAs senior leadership into the efficacy phase	IFS, AFC, LIIA, VRU
June-July 2023		Training of SPOCs and practitioners on evaluation requirements	IFS, AFC, LIIA, VRU
May-August 2023		Development of data collection interface	DXDigital with input from IFS, LAs, LIIA
June 2023		Development of grant agreements for LAs to participate in efficacy phase	VRU with support from LIIA, IFS
June-December 2023		Development of Data Sharing Agreement with a small number of LAs for sharing of historical data on young people allocated to teams participating in trial	LIIA, IFS, 3-4 LAs
June-July 2023		Development of Data Sharing Agreements supporting evaluation	IFS, AFC, LIIA
August 2023 – December 2024		Recruitment of young people and delivery of Your Choice or BAU support	LAs
August 2023 – December 2024		Collection of baseline data on efficacy phase participants	IFS, LAs
August 2023 – March 2025		Collection of process data on efficacy phase participants	IFS, AFC, LAs
December 2023 – March 2025		Collection of endline data on efficacy phase participants	IFS, LAs
November 2023 - March 2025		Collection of qualitative data	AFC

July – December 2024	Collection of cost data from Las	
August 2025	Final report on efficacy phase	

References

Barnao (2022). "The Good Lives Model: A Strength-Based Approach to Rehabilitating Offenders," *In* Garofalo, Sijtsema (eds) Clinical Forensic Psychology, Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-80882-2_27

Brooks, Lambert, Coulthard, Pennington, Kolehmainen (2021). "Social participation to support good mental health in neurodisability," *Child Care Health Dev*. 47: 675– 684, https://doi.org/10.1111/cch.12876

Butler, Chapman, Forman, Beck (2006). "The empirical status of cognitive-behavioral therapy: a review of meta-analyses," *Clin Psychol Rev.* 26(1):17-31, doi: 10.1016/j.cpr.2005.07.003

Case, S; Browning, A (2021). "Child First Justice: the research evidence-base [Full report]," Loughborough University. Report. https://hdl.handle.net/2134/14152040.v1

Crenna-Jennings, Hutchinson (2020). "Access to Child and Adolescent Mental Health Services in 2019," The Education Policy Institute, https://epi.org.uk/wp-content/uploads/2020/01/Access-to-CAMHS-in-2019_EPI.pdf

De Chaisemartin, C., & Ramirez-Cuellar, J. (2022). "At What Level Should One Cluster Standard Errors in Paired and Small-Strata Experiments?" Working Paper w27609, National Bureau of Economic Research, https://www.nber.org/papers/w27609Eels, T (1997). Handbook of Psychotherapy Case Formulation. New York: Guilford Press

Erikson (1968). Youth: Identity and crisis. New York, NY: Nortonhttp://dx.doi.org/10.1126/science.161.3838.257

Essau, Olaya, Anastassiou-Hadjicharalambous, Pauli, Gilvarry, Bray, O'callaghan, Ollendick (2012). "Psychometric properties of the Strength and Difficulties Questionnaire from five European countries," *Int J Methods Psychiatr Res.* 21(3): 232-45

Gaffney, H., Jolliffee, D., & White, H. (2021). "Trauma Informed Care: Toolkit technical report." London: Youth Endowment Fund.

Goodman, R. (1997). "The Strengths and Difficulties Questionnaire: a research note," *Journal of child psychology and psychiatry*, *38*(5), 581-586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., ... & Michie, S. (2014). "Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide," *BMJ 348*, doi: https://doi.org/10.1136/bmj.g1687

McConnell, B., & Vera-Hernández, M. (2015). "Going beyond simple sample size calculations: a practitioner's guide," IFS working paper, https://ifs.org.uk/sites/default/files/output_url_files/WP201517_update_Sep15.pdf

Ng Fat, L., Scholes, S., Boniface, S., Mindell J., & Stewart-Brown S. (2017). “Evaluating and establishing the national norms for mental well-being using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England,” *Quality of Life Research*, 26(5), 1129-1144

Quitangon (2019). “Vicarious trauma in clinicians: Fostering resilience and preventing burnout,” *Psychiatric Times*, 36(7), 18–19

Spaan, van den Boogert, Grootendorst-van Mil, Hoogendijk, Roza (2023). “Screening for disruptive behavior in adolescents at risk using the strengths and difficulties questionnaire,” *Journal of Research on Adolescence*, doi: https://doi.org/10.1111/jora.12858

Sweeney, Clement, Filson, Kennedy (2016). “Trauma-informed mental healthcare in the UK: what is it and how can we further its development?” *Mental Health Review Journal* 21(3), pp. 174-192

Steinberg (2008). “A Social Neuroscience Perspective on Adolescent Risk-Taking,” *Dev Rev.* 28(1):78-106. doi: 10.1016/j.dr.2007.08.002

Thomson, A., Dow, M.G., Kenardy, J.A., Johnston, D.W., Newman, M.G., Taylor, C.B (2007). “Prognostic indices with brief and standard CBT for panic disorder: II. Moderators of outcome,” *Psychological Medicine*, 37(10), pp.1503-1509.

Wenzel (2021). “The therapeutic relationship,” *In* A. Wenzel (Ed.), Handbook of cognitive behavioral therapy: Overview and approaches, pp. 175–203, American Psychological Association. https://doi.org/10.1037/0000218-007

Yao, Zhang, Zhu, Jing, McWhinnie, Abela (2009). “Measuring adolescent psychopathology: psychometric properties of the self-report Strengths and Difficulties Questionnaire in a sample of Chinese adolescents,” *Journal of Adolescent Health* 45, 55–62

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413