

Trajectories of female violent offending

Secondary Data Project Analysis Plan

Evaluating institutions: University of Bristol, University of Hull, University of the West of England

Investigator(s): Brennan I, Cornish RP, Hammerton G, Staines J, Teyhan A, Whelan M



Analysis Plan for YEF Secondary Data Analysis Projects

Project summary

Project title	Trajectories of female violent offending
Research Team	University of Bristol, University of Hull, University of the West of England
Principal investigator	Rosie Cornish
Analysis plan author(s)	Rosie Cornish, Alison Teyhan, Gemma Hammerton, Jo Staines, Michael Whelan, Iain Brennan
Overarching research question	What are the offending trajectories between the ages of 10 and 24 years among women and girls in England convicted or cautioned for violent offences?
Supporting research question(s)	Do these trajectories differ between ethnic groups, and across vulnerable subgroups such as those with care experience, educational needs and those experiencing deprivation? Is ethnicity an effect modifier in terms of the influence of these vulnerabilities on offending trajectories? What are the trajectories into violent offending, key transition points, and the potential impact of interventions on girls' behaviour from the perspectives of girls/young women, youth justice and probation professionals?
Dataset(s) to be used	The Ministry of Justice – Department for Education (MoJ-DfE) datashare; qualitative data – primary data collection.
Population characteristics	Individuals born between September 1995 and August 2000 attending state-funded schools in England
Years data spans	School data from academic years 2000/01 to 2011/12; police data from 2010 to 2024 ¹
Geographic coverage	England, UK
Primary outcome(s) investigated	All offending; violent offending
Main method(s) to be used or tested	Descriptive statistics to examine sample characteristics and patterns of offending

	Longitudinal latent class analysis to examine trajectories of offending Thematic analysis of interview and focus group data
--	--

Analysis plan history

Version	Date	Reason for revision
1.X [<i>latest</i>]		
1.1		
1.0 [<i>original</i>]		

Any changes to the design or methods need to be discussed with the YEF. Describe in the table above any agreed changes made to the design.

Table of contents

Trajectories of female violent offending.....	1
1. About the project.....	5
1.1. Background to the project.....	5
1.2. Research question(s).....	6
1.3. Hypotheses.....	7
1.4. Key concepts.....	8
2. About the datasets.....	8
2.1. Overview of datasets used	8
2.2. Secondary data source(s).....	9
2.3. Primary data collection.....	10
2.4. Linking datasets.....	11
2.5. Access and data protection.....	11
3. About the data	15
3.1. List of variables	15
3.2. Measurement of key concepts.....	16
3.3. Missing data and attrition	16
3.4. Other sources of bias.....	17
4. About the analysis	18
4.1. Overview of analytical approach	18
4.2. Approach to addressing research question(s).....	18
5. Project management	24
5.1. Risks and mitigations.....	24
5.2. Timeline	25
6. References.....	27

1. About the project

1.1. Background to the project

Previous research has indicated that there are two main trajectories of offending: life-course persistent, which starts around 10 years of age and persists into adulthood, and adolescent-limited, in which offending emerges in the mid-teens before quickly desisting (Moffitt and Caspi, 2001). More recent work (using the MoJ-DfE data share, the dataset we intend to use for this project) identified further subtypes, including some trajectories that only include non-violent offending and others which feature diverse offence types, including violence (Dickson, 2024). Most work examining offending trajectories has used either male-only or combined samples in which males dominate (Fontaine et al., 2009). However, there is evidence that females are less likely than males to be persistent offenders (Moffitt and Caspi, 2001; Dickson, 2024; Zheng and Cleveland, 2013) and some studies have identified trajectories specific to females. For example, a recent study among 15-64 year-olds found a female-specific trajectory in which offending peaked in mid-life (early 40s) (Hosseini 2025). Another study found two groups specific to females: a group of “early onset desisters” who tended to offend once or twice in early adolescence but then stopped, and an adult-onset group (Andersson et al., 2012). Finally, research conducted in a female only sample aged 10-31 years found four trajectories: life-course persistent, adolescent limited, young-adult limited, and adult onset (Coles et al., 2025). Having said this, other studies have identified quite similar overall trajectories, just with different prevalence among males and females (Caffman et al., 2015; Sivertsson et al., 2024). Other research has indicated that patterns of violent versus non-violent delinquency may differ among males and females following similar trajectories of overall offending, with a higher proportion of violent offending among males (Zheng and Cleveland, 2013). There is also evidence that males and females may have different risk factors for following different trajectories (Zheng and Cleveland, 2013) or differences in the extent of increased risk for a particular subgroup, such as those with care experience¹ (Hosseini, 2025).

It is well known that individuals with care experience and those from Black, Asian and Minority ethnic groups are over-represented in the criminal justice system in England and Wales (Laming, 2016; Lammy, 2017). In a group of 2.9 million children born between 1996 and 1999, 4% had received a conviction or caution by the age of 17 years; the corresponding rate was 33% among those with care experience (Hunter et al., 2023). Those from more deprived backgrounds and those with Special Educational Needs have also shown to be over-represented in the offending population (Department for Education and Ministry of Justice, 2022; Jahanshahi et al., 2022). The above work indicated that the extent of disproportionality

¹ Care experienced: someone who has lived (or is living) in care at some point during their childhood. This could be residential care, foster care, kinship care, or at home under a supervision order.

in terms of youth justice involvement among those with care experience varies according to ethnicity, with care experienced children having between 2- and 16-times higher rates of convictions and cautions compared to those without care experience, depending on their ethnicity (Hunter et al, 2023). Again, this work was carried out in a combined sample; it is not known whether ethnicity interacts with care experience and other factors in a different way in females compared to males or how these disproportionalities relate specifically to violent offending in females.

Violent offending is relatively rare in young people, particularly young females. In a population of just under 1.5 million young people in England, 1.7% of males and 0.3% of females had a conviction or caution for serious violence between the ages of 15 and 18 years (Cornish et al., 2022). This low prevalence of female violent offending in official records means that past studies have had to rely on small, self-selecting and homogeneous samples, which have limited generalisability. We are aware that work is currently underway (by a team with whom we are in contact) to examine overall offending trajectories in a female-only population using the MoJ-DfE data share. We plan to complement this work by focusing specifically on violence. We will examine trajectories of violent offending, both in its own right and in the context of overall offending among girls and young women who have been cautioned for or convicted of at least one violent offence. We will investigate how these patterns vary by ethnicity and explore the role of particular vulnerabilities captured in education data. Our focus is on females, but we will compare these trajectories to those among males.

The dominance of male subjects in violence research means that there is little theory of how females come to be involved in violence, how they desist from it, and whether male-focused interventions are suitable or effective for girls. For example, the dominant framework for explaining violence is the social-ecological framework (Bronfenbrenner, 1979) where factors like the effects of early childhood adversity combine with the influence of criminal peers, vulnerability to exploitation and incentives for offending in the local environment, as well as structural factors such as occupational opportunities and discrimination, to increase risk of involvement in violence (Billingham and Irwin-Rogers, 2022). However, boys and girls are exposed to and experience these factors in different ways (Haahr-Pedersen et al., 2020), meaning that how they contribute to violence is also likely to differ, as will the pathways to desistance (Widom and Osborn, 2021).

1.2. Research question(s)

Our research questions are:

1. What are the offending trajectories between the ages of 10 and 24 years among women and girls in England convicted or cautioned for violent offences?

2. Do these trajectories differ between ethnic groups, and across vulnerable subgroups such as those with care experience, educational needs and those experiencing deprivation?
3. Is ethnicity an effect modifier in terms of the influence of these vulnerabilities on offending trajectories?
4. How do the (i) offending trajectories and (ii) differences in these across ethnic groups and vulnerable subgroups observed in girls/young women compare to those seen among males?
5. What are the trajectories into violent offending, key transition points, and the potential impact of interventions on girls' behaviour from the perspectives of girls/young women, youth justice and probation professionals?

Table 1.2. How will the questions be addressed at each stage?

Question Number	Interim report	Final report
1, 4i	Our interim report will include an overview of the dataset and the different samples included in our analyses. We will present the results from the descriptive analysis examining female and male offending (overall and violence), as described in Table 4.1	Our final report will present methods and results for all research questions, adding to the descriptive results presented in the interim report (if necessary). We will interpret and discuss the implications of our findings.
2,3,4ii	We don't expect to have addressed these questions by this point.	
5	We will give an overview of progress to date in terms of sample recruitment and provide the interview schedule and focus group topic guide.	

1.3. Hypotheses

Based on findings from previous research, we expect to find one or more distinct trajectories of offending from those found in previous studies that have been male-only or combined male and female samples. Because we are considering a specific subset of females (i.e. those who have been convicted or cautioned for at least one violent offence) who, to our knowledge, have not previously been studied in terms of their offending trajectories, it is difficult to hypothesise the likely nature of these. We also expect there to be differences in

the prevalence of the trajectories across ethnic and other subgroups, with potentially higher prevalence of certain trajectories among those groups that are over-represented in the criminal justice system (some ethnic minority groups, care experienced, those with identified SEN, those who have experienced deprivation)

1.4. Key concepts

Table 1.4 Definitions of key concepts

Terms	Definition used
Offending	Having received any conviction or caution for any offence, excluding summary motoring offences
Violent offending	Any conviction or caution for an offence classified as violence against the person, possession of weapons, or robbery.
Female/male	This information is collected as part of the school census; allowed values are male and female. Since the academic year 2023/2024, this variable has been labelled as sex and the guidance states that this should be the sex of the pupil “as recognised in law” (Department for Education, 2025a). Prior to this, the variable was labelled as gender (with values male and female) (Department for Education, 2025b). Because guidance on collecting this information has changed over time, both in terms of the information being collected and who should provide this information, there will be inconsistencies over time in terms of what these labels mean. Where we need to refer to this variable, we will use the term “sex/gender” but will avoid this wherever possible, using the terms female and male instead. This will refer to the information as recorded in the NPD.
Ethnicity	Information on ethnicity is also collected as part of the school census. Current guidance states that this information must be provided by the parent or guardian and not ascribed by the school (Department for Education, 2025a). However, as with sex/gender, guidance has changed over time and, previously, ethnicity could be ascribed by the pupil’s school. Out of necessity, we can only use the information as provided in the NPD.

2. About the datasets

2.1. Overview of datasets used

This project will use the MoJ-DfE data share, a linkage between the National Pupil Database (NPD) and the Police National Computer (PNC) for individuals born after 31st August 1985. The NPD is a central repository of education data for children attending state schools in England. The PNC is a database that contains information on convictions and cautions for recordable offences in England and Wales.

2.2. Secondary data source(s)

Table 2.2a Dataset Description

Name of dataset	MoJ-DfE datashare
Data owner(s)	Ministry of Justice, Department for Education
Type of data	Administrative
Availability of data	Data must be applied for (https://www.gov.uk/government/publications/moj-data-first-application-form-for-secure-access-to-data); data access is via the Office for National Statistics (ONS) Integrated Data Service (IDS) or Secure Research Service (SRS), a trusted research environment
Team member(s) who will have access	Alison Teyhan, Rosie Cornish, Gemma Hammerton
Population/geographic coverage or sampling frame	Individuals attending state-funded schools in England
Years covered or survey waves	We will include individuals born between September 1995 and August 2000; offending data will cover the period up to December 2024.
Exclusion criteria	Individuals with unknown male/female status (estimated to be 0.03% of the population)
Expected population/sample size (following exclusion criteria)	3.5 million for the initial descriptive analysis (50% female). We estimate that between 8,750 and 13,125 females will have a conviction or caution for violence (and expect there to be approximately five times as many males with a conviction or caution for violence). The numbers with a conviction or caution for any offence will be higher.
Documentation	https://assets.publishing.service.gov.uk/media/683863d5c99c4f37ab4e867c/Data_First_User_Guide_Version_8.2.pdf

2.3. Primary data collection

Qualitative data will be collected to address our fifth research question². This section provides an overview of the qualitative component of the research, exploring the perceptions of two key stakeholder groups: probation and community youth workers working with girls/young women, and girls/young women convicted or cautioned for violent offences. Two qualitative research methods will be used – focus groups and interviews – which are discussed in more detail below. The focus groups will be run with practitioners, while the interviews will be run with the girls/young women. The qualitative research process can be divided into four key phases, as outlined below.

Phase one: planning and preparation

In this phase, the qualitative members of the research team (MW, JS, IB) will ensure that mechanisms are in place to begin the data collection process. This will include an application for ethical approval and a steering group meeting. Emerging themes and insights from the quantitative data analysis will provide an orientation for the themes explored within the focus groups and interviews.

Phase two: data collection and analysis 1 – focus groups

Two focus groups will be held with youth justice, probation and community youth workers (total n=10) working with girls/young women, contacted through the research team's contacts. The team will attempt to ensure a geographic spread in focus group participants, as well as a variation in practice contexts. The focus groups will be held online, will last for approximately two hours and will be informed by a short topic guide, to explore professional perspectives on girls' trajectories into violent offending, key transitions and turning points, and the impact of interventions on girls' behaviour. Professionals will be asked to discuss their perspectives of the impact of intersectional characteristics, such as ethnicity, care experience and educational needs, on girls' trajectories. Insights from these discussions will feed into more in-depth and specific discussions within the second stage of data collection (interviews with girls/young women).

Phase three: data collection and analysis 2 – interviews

Interviews will be conducted with girls/young women convicted or cautioned for violent offences (n=10). The girls/young women will be contacted via an existing trial (YEF Another Chance Focused Deterrence multicentre trial) and invited to participate in an individual online narrative interview to explore their understanding of how and why they became involved in

² What are the trajectories into violent offending, key transition points, and the potential impact of interventions on girls' behaviour from the perspectives of girls/young women, youth justice and probation professionals?

violent offending behaviour, the impact of any interventions or support they may have received, and key points at which they believe their pathways into offending behaviour and the criminal justice system could have been diverted. A purposive sampling approach will be adopted, and attempts will be made to include participants with as wide a range of demographic characteristics as possible, to enable consideration of intersectionality. Issues of potential discrimination on the basis of different characteristics will be explored with all interview participants. A semi-structured interview approach will be adopted, allowing a level of flexibility for participants to highlight areas that might not have been expected, whilst also retaining an orientation to the core research questions. The questions will explore life course periods that overlap with the exposure periods in the project's quantitative analysis, allowing us to contextualise the findings from the quantitative data analysis. In addition, we will also focus on more recent exposures, such as exploitation, that are not available in the quantitative data, thus providing greater insight into trajectories into offending for this population.

The girls and young women will be paid for their involvement in the interview (£15 per hour in Love2Shop vouchers), in recognition of the time they give to the study.

2.4. Linking datasets

Quantitative data: not applicable - linkage between the NPD and PNC has already been carried out by the Department for Education and Ministry of Justice.

Qualitative data: not applicable – the qualitative data will not be linked to any other data.

2.5. Access and data protection

Quantitative data

1. Data Protection

Data access is granted via the ONS IDS/SRS. All researchers accessing data must adhere to the principles of the ONS 'Five Safes framework' (safe data, safe projects, safe people, safe settings, and safe outputs). Ethical approval for the work has been granted by the University of Bristol Faculty of Health Sciences Ethics Committee.

Safe Data

- (1) the dataset will be minimised to only include variables necessary for the project (See Table 3.1);
- (2) the data will have been de-identified prior to access;

Safe projects

This project was approved by the Ministry of Justice and Department for Education in October 2025.

Safe People

Data access will only be granted to individuals who are ONS accredited researchers, who have been DBS checked (basic) and are named on the data application (see Table 2.2a).

Safe setting

The data will be stored on project-specific folders on the ONS IDS or SRS. This is a trusted research environment. Access is via an ONS approved SafePod, Safe Room, or an office with Assured Organisational Connectivity, using multi-factor authentication. Access to the data will cease 12 months after the end of the project, on 31/08/2028.

Safe Outputs

All outputs will be disclosure checked by the ONS prior to release from the IDS/SRS. Because we will be using Child Looked After (CLA) data, all our figures will be rounded to the nearest 10 and percentages to the nearest whole number, and any percentage that round to zero will be suppressed.

Qualitative data

Data will be generated prospectively as part of the project under conditions of informed consent for the team to analyse the data and to use excerpts of the data in project outputs.

Safe Data

At the time of collection, the interview data will contain sensitive and personally-identifiable data. The focus groups are less likely to generate sensitive data, but they are likely to include personally-identifiable data. Qualitative interviews will be collected in person and recorded using a digital voice recorder.

Focus group interviews will be conducted remotely and will be recorded using Microsoft Teams. The recording will be automatically saved to the interviewer's password-protected OneDrive for Business account, which is Advanced Encryption Standard (AES) encrypted.

Safe projects

The project will obtain ethical approval from the University of Hull.

Safe People

Interviews and focus groups will be conducted by a researcher with an Enhanced DBS check. All qualitative data will be collected and processed under informed consent. The role and/or organisation (i.e. University of Hull research team; TP Transcription) of individuals who will have access to unprocessed (i.e. not yet anonymised) data will be clearly specified in the consent process. Participants will be asked to opt in to their data being passed to TP Transcription for transcription. If they choose not to opt in, transcription will be carried out by one of the University of Hull research team.

All University of Hull staff handling the unprocessed data within the Data Safe Haven are required to completed NHS Digital Security Awareness training as well as routine Data Protection training.

Safe setting

Interview data will be collected using a digital voice recorder. Immediately following interview, the digital file will be uploaded to a local encrypted drive on a password-protected and Cyber Essentials Plus compliant laptop. The audio file will then be deleted from the voice recorder.

The audio files will be transferred to a third-party transcription service (TP Transcription) via the service's HTTPS/SFTP TLS 1.3 with AES-256 encrypted portal. This processing is governed by a data sharing agreement with University of Hull.

The transcribed file will be returned to University of Hull and immediately transferred to the University of Hull trusted research environment (Data Safe Haven) and deleted from local storage. The Data Safe Haven facility is NHS Digital Toolkit/ISO227001 compliant for the safe handling, processing and storage of sensitive data, is fully auditable and is managed by a full-time data manager. The unprocessed data and audio recordings will be deleted once the files have been anonymised and the Data Safe Haven project will be torn down, during which the data will be electronically destroyed (ISO27001) and a certification of destruction will be issued.

The unprocessed files will be anonymised while in the Data Safe Haven following UK Data Service/UKRI guidance before being egressed. The anonymised files will be stored locally, backed up on a University of Hull Microsoft Teams project channel and uploaded to NVivo cloud facility for analysis. The anonymised data will be stored on the Teams channel for up to five years, after which it will be electronically deleted.

Safe Outputs

All quotes will be considered at risk of statistical disclosure. Accordingly, they will be assessed for risk of identifiability disclosure by two researchers independently using UK Data Service guidance.

Participants will be asked to choose a pseudonym that is not a nickname or disclosive of their identity. The researchers will reserve the right to change this pseudonym if they have reason to believe that it is disclosive.

2. Data Processing Roles

Quantitative data

The Data Controllers are the Ministry of Justice and Department for Education; the approved researchers are Data Processors.

Qualitative data

University of Hull will be the data controller.

3. Legal Basis for data processing

Quantitative data

The legal basis for using participants' information, under GDPR and the Data Protection Act 2018, is: Public task (article 6 (1)(e): The processing is necessary for us to perform a task in the public interest or for our official functions, and the task or function has a clear basis in law.

Our justification for this is as follows: the University of Bristol is founded and regulated by its Charter and related Acts of Parliament. Together they define the University's role as the provision of teaching and research. All research falls within this lawful basis.

We believe this project is in the public interest because it has the potential to improve understanding of which girls and young women are most likely to get involved in violent offending. Without this understanding, interventions are likely to be aimed at the wrong individuals. Efforts to reduce offending and re-offending of national social and economic benefit.

Qualitative data

The legal basis for using participants' information, under GDPR and the Data Protection Act 2018, is: Informed consent (article 6 (1)(a): The data subject has given consent to the processing of his or her personal data for one or more specific purposes.

We will seek written informed consent from all participants with an opt-in requirement for the unprocessed audio recording of the interview to be processed by a third-party transcription provider.

4. Data access process

Quantitative data

Anyone wanting to access the MoJ-DfE must apply to the MoJ and DfE, specifying which datasets and variables they would like to access and for which years of coverage. Applications are reviewed by both the MoJ (by their Data Access Group and, subsequently, the Data Access Governance Board, chaired by the MoJ's chief statistician) and the DfE (by the Data Sharing Approval Panel). On approval, researchers are provided with a data sharing agreement and schedule to sign. A dataset is then built for the project and deposited within project-specific folders within the IDS/SRS. The dataset can only be accessed by approved members of the project team via a SafePod, Safe Room or an office with Assured Organisational Connectivity.

Qualitative data

The anonymised qualitative data will be accessible by all research team members on the NVivo cloud platform.

3. About the data

As these items do not readily apply to qualitative data, section 3 refers exclusively to the quantitative data.

3.1. List of variables

Table 3.1: Variable definitions

Variable abbreviation	Variable definition	Variable source	Derivation or specification
Outcomes			
Violence	Any conviction or caution for a violent offence	PNC	Any offence classified as violence against the person, weapon carrying or robbery
Offending	Any conviction or caution for any offence, excluding summary motoring offences	PNC	All offences, except those classified as summary motoring offences
Exposures			
Age	Age at offence	NPD & PNC	Will be derived from month and year of birth recorded in the NPD and date of offence recorded in the PNC
Care experienced	Ever been looked after up to age 10	NPD (CLA, all episodes)	Derived from episodes of care data
Experienced deprivation	Ever been eligible for free school meals (FSM) or had IDACI ¹ score in the top decile up to age 10	NPD (school census data)	Derived from FSM eligibility flags and IDACI scores
SEN	Any recorded SEN status in primary school	NPD (school census data)	Categories: No SEN, SEN support, EHC plan
Ethnicity	Recorded ethnicity in primary school	NPD (school census data)	We will use as granular categories as possible (see Table 3.2). Small numbers in some categories may require us to combine groups.

Other variables			
Local authority	Local authority of the school	NPD (school census data)	As recorded in the NPD, one of 317

1. IDACI: Income Deprivation Index Affecting Children, a measure defined at Lower Layer Super Output Area (LSOA) level, derived from a pupil's home postcode, indicating the proportion of children aged 0-15 years living in income-deprived families in that area (so higher scores indicate greater levels of deprivation).

3.2. Measurement of key concepts

Table 3.2 Measurement of key concepts

Concept	How the concept will be measured and encoded
Female/male	See Table 1.4 for a description of sex/gender recording in the NPD.
Ethnicity	See Table 1.4 for a description of how ethnicity is recorded in the NPD. Various ethnicity categories have been used over the years. Currently the NPD contains three ethnicity variables, one with broad categories, termed "major" ethnic groups, one with more granular categories, termed "minor" ethnic groups, and one with even more granular categories. We would like to use minor ethnic group if possible (the more granular variable is likely to contain too many subgroups with very small numbers of pupils), but acknowledge that some groups may need to be combined if numbers are too small. Such decisions will be partly data driven to retain as much disaggregation as possible, but we will only aggregate within the major ethnic groups. Major ethnic groups: Asian, Black, Chinese, Mixed, White, any other ethnic group, Unclassified Minor ethnic groups: Bangladeshi, Indian, any other Asian background, Pakistani, Black African, Black Caribbean, any other Black background, Chinese, White and Asian, White and Black African, White and Black Caribbean, any other mixed background, White British, White Irish, Traveller of Irish heritage, Gypsy/Roma, any other White background, any other ethnic group (other possible recorded values are: refused, invalid, missing, not applicable, and information not yet obtained)

3.3. Missing data and attrition

(i) Key variables that you'd ideally access to address the research question, but will not be present in the datasets

Experience of violence (as a victim), criminal exploitation or coercion.

The dataset does not capture this type of information, but we plan to explore these issues as part of the qualitative element of our project (see section 2.3).

(ii) Observation level data that is incomplete for a subset of the population being studied

Based on previous work, we expect male/female status to be recorded for over 99% of pupils. Ethnicity, FSM status, IDACI scores and SEN status are likely to be missing (or refused/unclassified in the case of ethnicity) for a small percentage of pupils.

(iii) Potential approaches to mitigating the impact of missing data

We will define our sample as those who appear in at least one school census during key stage 1 (aged 4-7), at least one during key stage 2 (aged 7-11) and have key stage 4 attainment results (obtained at age 16). Thus, we will not require individuals to have appeared in every school census during primary school in order to derive exposure status. At the same time, this approach will ensure that those in our sample were still in the English education system at age 16. This will minimise the amount of missing data. However, as stated above, we do not think that missing data will be a major issue in this dataset.

3.4. Other sources of bias

The school data only include those attending state schools in England. Therefore, children who attend private schools, are home-schooled, or are not registered at a state school in England for any other reason, will not be in our sample.

There are several potential sources of bias with regards the PNC data. In brief, there is known to be bias in terms of whose criminal behaviour is detected by the police, and the disposal type they are given. An example of this is variations in the rate of reporting of crime across communities and demographic groups. Bias may also be introduced through the data linkage process if specific subgroups of the population are less likely to be matched or more likely to be mis-matched. This would be more likely to happen in cases where identifier information (e.g. name, date of birth) held in the NPD and PNC were poorly recorded over time. In addition, the PNC data included in the dataset primarily includes offences committed in England and Wales. Thus, individuals who move permanently out of England and Wales after age 16 may have inaccurate offending histories. Given that some groups may be more or less likely to move, this could result in differential measurement error. However, we expect this to apply to a relatively small proportion of individuals.

Research has indicated that many young female perpetrators of violence are / have been victims of criminal exploitation and/or coercion by partners. These proximal factors (i.e. close in time to the offence) will not be captured in the dataset as the PNC only includes information about perpetration and these exposures will not appear in education records. We will acknowledge these limitations and biases in the interpretation of our results. We plan to explore these issues in the qualitative element of the project.

4. About the analysis

4.1. Overview of analytical approach

Quantitative data

Our descriptive analyses will include an overview of the dataset, the derivation of our specific study samples for each research question, and a summary of each of our exposures, including details of any missing data. We will then describe the prevalence of offending (overall and violent) at different ages, the average age at first offence, and types of offence at different ages, both among individuals with at least one conviction or caution for violence and among those with no convictions or cautions for violence. Further details are provided in Table 4.1.

We will then use longitudinal latent class analysis (Herle et al., 2020) to derive offending trajectories for violent and non-violent offences, accounting for clustering of individuals within local authorities using robust standard errors. Using a bias-adjusted three-step method (Vermunt, 2010), we will use multinomial logistic regression to examine trajectory membership across ethnic groups and according to key vulnerabilities (care experience, SEN, deprivation). The three steps referred to in this approach are: 1) building the latent class model, 2) assigning individuals to these classes based on their posterior probabilities of class membership, 3) using multinomial logistic regression to examine associations between covariates and these assigned classes. It has been shown that use of a three-step approach (as opposed to simultaneously modelling the latent classes and the covariate associations) underestimates the associations between covariates and class membership. The bias-adjusted three step methods involves correcting for the classification errors inherent in the second step by fixing the class-specific intercepts at values corresponding to the estimated error rates when fitting the multinomial logistic regression model at step three. We will investigate whether there are interactions between ethnicity and these vulnerabilities. Further details are provided in Table 4.1.

Qualitative data

Data from the interviews and focus groups will be analysed by thematic analysis (Braun and Clarke, 2008).

4.2. Approach to addressing research question(s)

Table 4.1: Research question(s): approach and methods

Research question	[1 and 4i]: What are the offending trajectories between the ages of 10 and 24 years among women and girls in England convicted or cautioned for violent offences? How do the offending trajectories observed in girls/young women compare to those seen among males?
Hypothesis, if relevant	N/A

<i>What will you be able to say by the interim report</i>	What is the prevalence of overall and violent offending from ages 10 to 24 among (i) females and (ii) males? What are the patterns of offending (age at first offence, number and types of offence at different ages) among females and males who have at least one conviction or caution for violence by age 25?
<i>Descriptive analysis, if relevant</i>	We will first look at the prevalence of overall female offending and violent offending by age (from 10 to 24 years). We will then restrict the sample to those with at least one conviction or caution and divide this group into those who have a conviction or caution for at least one violence offence before age 24 and those who have no convictions or cautions for violence. We will describe offending characteristics of these two groups: age at first offence; number of offences between 10-13, 14-17, 18-20 and 21-23; and types of offending. We will carry out a similar descriptive analysis among males. After deriving the trajectories (see below), we will describe these in terms of average age at first offence and first violent offence, frequency and type of offending at different ages, and age at desistance (where applicable).
<i>Models, specifications and statistical techniques used, if relevant</i>	We will use longitudinal latent class analysis to derive offending trajectories for violent and non-violent offences using repeated measures of offending across the age range. These measures will either be binary (no offence vs at least one offence) or categorical (no offence, one offence, two or more offences); this will depend on the distribution of these repeated measures. To examine whether the trajectories of violent/non-violent offending are comparable for females and males, we will test for measurement invariance using stepwise multiple indicator multiple cause modelling (Masyn, 2017). This method allows the prevalence of each identified trajectory to differ across females and males and examines whether constraints can be made so the number and shape of the trajectories is the same.
<i>Estimating equation, if relevant</i>	N/A

<i>What does the approach need to succeed (constraints/assumptions)?</i>	If the derived trajectories are all parallel and/or fit statistics do not point to a clear class solution, this approach may not be useful. However, based on previous research, we feel this is unlikely.
<i>Uncertainty and inference</i>	Model fit will be assessed using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and sample-size adjusted Bayesian Information Criterion (aBIC), with lower values indicating better model fit. We will also use likelihood ratio tests (Bootstrap Likelihood Ratio Test; BLRT and Lo, Mendell & Rubin Likelihood Ratio Test; LMR-LRT) to assess the improvement in model fit for each additional class (with a high p value indicating no evidence of improvement in fit from model with one less class).
<i>Robustness checks</i>	Assignment probabilities will be used to examine classification accuracy and class separation will be assessed using entropy. If entropy is between 0.6 and 0.7, we will explore using the one-step (as opposed to the three-step) approach to examine associations between vulnerability factors and classes, as this has been shown to work well even with low entropy. If entropy is below 0.6, we will examine population-level changes in offending rather than using latent classes.
<i>Subgroup you intend to study</i>	Females (although a comparative analysis will be carried out on males)
<i>Changes to the analysis</i>	Unlikely

<i>Research question</i>	[2, 3 and 4ii]: Do the offending trajectories differ between ethnic groups, and across vulnerable subgroups such as those with care experience, educational needs and those experiencing deprivation? How do the differences in these across ethnic groups and vulnerable subgroups observed in girls/young women compare to those seen among males? Is ethnicity an effect modifier in terms of the influence of these vulnerabilities on offending trajectories in females?
<i>Hypothesis, if relevant</i>	N/A

<i>What will you be able to say by the interim report</i>	We do not expect to have started this work before the interim report. However, we will be able to report on sample size for this analysis.
<i>Descriptive analysis, if relevant</i>	N/A
<i>Models, specifications and statistical techniques used, if relevant</i>	Using a bias-adjusted three-step method, we will use multinomial logistic regression to examine trajectory membership across ethnic groups and according to key vulnerabilities (care experience, SEN, deprivation). We will investigate whether there are interactions between ethnicity and these vulnerabilities. If our analysis (detailed above, objective 4i) suggests that the identified trajectories are comparable for females and males, we will examine whether any differences in the trajectories across ethnic groups and vulnerable subgroups are similar among females and males.
<i>Estimating equation, if relevant</i>	N/A
<i>What does the approach need to succeed (constraints/assumptions)?</i>	While our sample size for females (our main group of interest) will be large enough to examine overall trajectories and patterns of offending, as well as those in key subgroups, it may not be sufficient to consider some of the smaller ethnic groups individually or to examine all aspects of intersectionality.
<i>Uncertainty and inference</i>	Results from the multinomial logistic regression (examining associations between characteristics and trajectory membership) will be presented as odds ratios with 95% confidence intervals.
<i>Robustness checks</i>	
<i>Subgroup you intend to study</i>	Individuals with at least one conviction or caution for violence between 10 and 24 years.
<i>Changes to the analysis</i>	If the trajectories for males and females are too different to compare the strength of association for risk factors (e.g. using interaction), we will report results for females and males separately but will not statistically compare them.

<i>Research question</i>	[5]: What are the trajectories into violent offending, key transition points, and the potential impact of interventions on girls' behaviour from the perspectives of girls/young women, youth justice and probation professionals?
<i>Hypothesis, if relevant</i>	The analysis is interpretivist, thus does not seek to test a hypothesis. The accounts of the respondents will be triangulated with the results of the quantitative analysis to provide context and to provide nuanced, but sampling-limited insights that are not available from the quantitative data.
<i>What will you be able to say by the interim report</i>	We do not expect to have started this work before the interim report. However, we will be able to report on participant recruitment and provide our interview schedule and topic guide for the focus groups.
<i>Descriptive analysis, if relevant</i>	Data from the interviews and focus groups will be analysed by thematic analysis.
<i>Models, specifications and statistical techniques used, if relevant</i>	N/A
<i>Estimating equation, if relevant</i>	N/A
<i>What does the approach need to succeed (constraints/assumptions)?</i>	A key vulnerability in this approach is the ability to recruit both practitioners and young women. As a research team we have a broad base of contacts upon which we will draw. In particular, we plan to use an existing project (YEF Another Chance Focused Deterrence multicentred trial) to support recruitment. The available lead-in time to this phase of the research will be valuable in ensuring we can open conversations with potential participants from an early point.
<i>Uncertainty and inference</i>	N/A
<i>Robustness checks</i>	Triangulation of the quantitative and qualitative data sources provides an informal form of bidirectional robustness checking.
<i>Subgroup you intend to study</i>	Two key stakeholder groups: probation and community youth workers working with girls/young women (total

	n=10), and girls/young women convicted /cautioned for violent offences (total n=10)
<i>Changes to the analysis</i>	N/A

5. Project management

5.1. Risks and mitigations

Table 5.1 Risks and mitigations

Number	Risk	Likelihood (Low/Medium/ High)	Mitigation
1	Data access delays	Medium	Data application submitted in mid-June
2	Data access issues	Medium	Data access issues (during the project – e.g. due to ONS maintenance activities) are likely to be short term (time-limited). We will minimise their effect by carrying out alternative activities until access is restored – e.g. reviewing the literature, writing elements of the report(s).
3	Failure to recruit sufficient girls/young women	Medium	Extend the invitation to participate to girls/young women involved with charities and other specialist organisations working with girls/young women involved in violence
4	Delays in installing Mplus (software) within the SRS	Low	Initiate the process as soon as we have data access (even though we don't need the software for the first objective).
5	Data breach or deductive disclosure	Low	Quantitative data: The quantitative data will be stored in the ONS SRS and there are strict rules governing access. Although there is a theoretical risk of a data breach, we feel this risk is very low. The ONS carry out disclosure checks before any results are exported. AT, RC and GH (the only members of the team who will have access to the data) are ONS accredited researchers and understand their responsibilities regarding data protection and security. For the qualitative data, we will follow institutional policies regarding data governance, as outlined above.

6	Delays approving outputs for export	Medium	Ensure outputs are clearly annotated as required and that all statistical disclosure control rules have been adhered to. Submit outputs for approval with enough time to allow for potential delays (i.e. to allow reports to still be submitted on schedule).
7	Number of participants within specific subgroups is small	Medium	We will aggregate ethnic and SEN groups (could be dichotomised into SEN/No SEN) if needed and may not be able to investigate all aspects of intersectionality.

5.2. Timeline

Table 5.2 Timeline

Date	Activity	Staff responsible/leading
August 2025	Analysis protocol submission	RC
August - December 2025	Data application, ethics approval for quantitative analysis	RC
January – March 2026	Planning and recruitment for interviews and focus groups	JS, MW, IB
March – April 2026	Data cleaning, initial descriptive analysis	AT (RC advising)
30 th April 2026	Interim first draft report to YEF	AT (with input from all)
May 2026 – March 2027	Main analysis	AT (with GH, RC and IB advising)
15 th June 2026	Interim report final draft to YEF	AT (with input from all)
September - December 2026	Conduct interviews and focus groups	JS, MW, researcher
January - March 2027	Qualitative data analysis	JS, MW, researcher
March – June 2027	Report and paper writing	AT (with input from all)

15 th July 2027	Final report first draft to YEF	AT; JS/MW – qualitative work (with input from all)
31 st August 2027	Final report final version to YEF	AT (with input from all)

6. References

- Andersson F, Levander S, Svensson R, Levander MT. Sex differences in offending trajectories in a Swedish cohort. *Crim Behav Ment Health* 2012; 22:108-121.
- Billingham L and Irwin-Rogers K. *Against youth violence - a social harm perspective*. Bristol University Press; 2022.
- Braun V and Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006; 3(2), 77-101.
- Bronfenbrenner U. *Ecology of human development*. Harvard University Press; 1979.
- Cauffman E, Monahan KC, Thomas AG. Pathways to persistence: female offending from 14 to 25. *J Dev Life Course Criminol* 2015; 1:236-68.
- Cornish R, Tilling K, Brennan I. Ministry of Justice – Department for Education linked dataset. Feasibility of evaluating early interventions for violence prevention: generating matched control groups. ADR UK, 2022. [Available from: https://www.adruk.org/fileadmin/uploads/adruk/Documents/Feasibility_study_2_MoJ-DfE_linked_dataset_Generating_matched_controls_.pdf; accessed 28th January 2026].
- Department for Education and Ministry of Justice 2022. Education, children’s social care and offending. Descriptive statistics. [Available from: https://assets.publishing.service.gov.uk/media/6227a9b58fa8f526dcf89e17/Education_children_s_social_care_and_offending_descriptive_stats_FINAL.pdf; accessed 28th January 2026].
- Department for Education 2025a. Complete the school census. Data items 2024/2025. [Available from: <https://www.gov.uk/guidance/complete-the-school-census/data-items-2024-to-2025>; accessed July 2nd 2025].
- Department for Education 2025b. Schools, pupils and their characteristics. Academic year 2024/25. [Available from: <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics/data-guidance>; accessed July 2nd 2025].
- Dickson H. Childhood educational predictors of re-offending trajectories. Data Insight Report. ADR UK; 2024.
- Fontaine N et al. Research review: A critical review of studies on the developmental trajectories of antisocial behavior in females. *J Child Psychol Psychiatr* 2009; 50:363-385.
- Haahr-Pedersen I et al. Females have more complex patterns of childhood adversity: implications for mental, social, and emotional outcomes in adulthood. *Eur J Psychotrauma* 2020; 11(1):1708618.
- Herle M et al. Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *Eur J Epidemiol* 2020; 35:205-22.
- Hossein G. A prospective examination of offending trajectories in men and women with out-of-home care experience followed up until retirement age. *Child Abuse Neglect* 2025; 161:107253.
- Hunter K, Francis B, Fitzpatrick C. Care experience, ethnicity and youth justice involvement: key trends and policy implications. Policy Briefing. ADR UK; 2023.

- Jahanshahi B, Murray K, McVie S. ACEs, places and inequality: understanding the effects of adverse childhood experiences and poverty on offending in childhood. *Br J Criminol* 2022; 62(3):751-72.
- Laming H. In care, out of trouble. Prison Reform Trust; 2016.
- Lammy H. An independent review into the treatment of, and outcomes for, Black, Asian, and Minority Ethnic individuals in the criminal justice system. HM Government; 2017.
- Masyn KE. Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling* 2017; 24(2):180-97.
- Moffitt TE and Caspi A. Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Dev Psychopathol* 2001; 13(2):355-375.
- Sivertsson F, Carlsson C, Almquist YB, Brännström L. Offending trajectories from childhood to retirement age: findings from the Stockholm birth cohort study. *J Crim Justice* 2024; 91:102155.
- Vermunt JK. Latent class modelling with covariates: two improved three-step approaches. *Political Analysis* 2010; 18:450-469.
- Widom CS and Osborn M. The cycle of violence: abused and neglected girls to adult female offenders. *Forensic Criminol* 2021; 16(3):266-85.
- Zheng Y and Cleveland HH. Identifying gender-specific developmental trajectories of nonviolent and violent delinquency from adolescence to young adulthood. *J Adolesc.* 2013; 36(2):371-81.