

# An examination of the association between school absence and exclusion and violent crime in the ALSPAC cohort



## Secondary Data Project Analysis Plan

Evaluating institution: University of Bristol

Principal investigator(s): Teyhan A, Cornish RP, Brennan, I

## Analysis Plan for YEF Secondary Data Analysis Projects

### Project summary

<b>Project title<sup>1</sup></b>	An examination of the association between school absence and exclusion and violent crime in the ALSPAC cohort
<b>Research Team</b>	University of Bristol and University of Hull
<b>Principal investigator</b>	Alison Teyhan
<b>Analysis plan author(s)</b>	Alison Teyhan, Rosie Cornish, Iain Brennan
<b>Overarching research question<sup>2</sup></b>	What is the association between: (1) school absence and (2) school exclusion and: (a) self-reported and (b) official sanction for violent behaviour.
<b>Supporting research question(s)<sup>3</sup></b>	
<b>Dataset(s) to be used</b>	- Avon Longitudinal Study of Parents and Children (ALSPAC) - Avon & Somerset Police (A&SP) Data - National Pupil Database (NPD)
<b>Population characteristics</b>	Our sample will be all individuals in ALSPAC for whom we have permission to link to crime and education data. We will exclude the small number who have a police record

<sup>1</sup> Please make sure the title matches what's in the header.

<sup>2</sup> In simple terms written for a none-expert, what's the main thing this research projects sets out to answer?

<sup>3</sup> What are the supporting research questions that will be tested in support of addressing the primary research question. This should not exceed three on the cover sheet and more detail can be added below if there are further questions to be addressed.

	relating to serious violence prior to the start of Key Stage 4 (KS4, school years 10 and 11, age 15-16 years).
<b>Years data spans</b>	ALSPAC - participants born in 1991/92. A&SP data – available from 2007 to 2021 (small number of pre-2007 records). NPD – absence and exclusion records available for the years the cohort were in KS4 (2005-2009).
<b>Geographic coverage</b>	Avon and Somerset, UK
<b>Primary outcome(s) investigated</b>	(1) Police record for serious violence aged 16-18yrs (2) Self-reported serious violence aged 17-20yrs
<b>Main method(s) to be used or tested</b>	To model the relationship between our exposures and outcomes we will use multilevel regression models to account for clustering of the sample within schools. This will be Poisson regression (for police-recorded violence) and logistic regression (for self-reported violence). Models will be adjusted for age and other confounders.

### Analysis plan history

Version	Date	Reason for revision
1.1	21st Feb 2024	Incorporating changes suggested by YEF
1.0 [original]	7 <sup>th</sup> Nov 2023	

*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

<b>1. About the project</b>	<b>4</b>
1.1. Background to the project	4
1.2. Research question(s)	5
1.3. Hypotheses	6
1.4. Key concepts	7
<b>2. About the datasets</b>	<b>8</b>
2.1. Overview of datasets used	8
2.2. Secondary data source(s)	8
2.3. Primary data collection	11
2.4. Linking datasets	11
2.5. Access and data protection	12
<b>3. About the data</b>	<b>14</b>
3.1. List of variables	14
3.2. Measurement of key concepts	20
3.3. Missing data and attrition	21
3.4. Other sources of bias	22
<b>4. About the analysis</b>	<b>22</b>
4.1. Overview of analytical approach	22
4.2. Approach to addressing research question(s)	23
<b>5. Project management</b>	<b>26</b>
5.1. Risks and mitigations	26
5.2. Timeline	27
<b>6. References</b>	<b>28</b>

## 1. About the project

### 1.1. Background to the project

Academic performance, special educational needs and school experience are strongly correlated with later criminal justice outcomes. Strong academic performance is a protective factor against violent behaviour, while underperformance, special educational needs, exclusion, persistent absence, having under-performing or delinquent peers and area-level inequalities are correlated with violence and involvement with the criminal justice system. Of these risk factors for serious violence, absence and exclusions are of particular interest because they are modifiable (e.g. via changes in school actions and policy - as observed in Scotland - or through school-based interventions).

Understanding the relationship between absence from school – unenforced, in the form of absence, or enforced, in the form of suspension or exclusion – and violence is a pressing issue. There are strong beliefs among some policy-makers and researchers that exclusions, in particular, are a direct cause of later offending and, on this basis, some school areas have moved to prohibit exclusions. While there are many harms associated with absence or exclusion, misconstruing the nature of the link between them and violence risks stigmatising those children and distracting attention from the underlying causes of violence. We seek to estimate the connections between absence/exclusion and violence in a way that will better inform school policies and also better inform the prioritisation of violence prevention interventions with young people.

#### **Absence – a sign of underlying issues or a turning point for violence?**

For many children who go on to be involved in violence, persistent absence is a common feature of their school record. In the 2012/13-2014/15 cohort of Key Stage 4 children, the relative risk of subsequently being involved in serious violence was 1.93 (95% CI 1.92-1.94) for those who were persistently absent compared to those who were not. When the reason for that absence was unexplained, which would include truancy, this relative risk rose to 7.00 (95% CI 6.89-7.11). However, it is unclear if this is a causal relationship. It could be that absence and violence have a common cause, rather than absence being a turning point for involvement in violence.

#### **Exclusion and violence are correlated but the causal evidence is weak**

Similarly, exclusion from school, either temporarily or permanently, is a relatively common event in the lives of teenagers convicted of serious violence: 15% of individuals who had been convicted or cautioned for a serious violent offence were permanently excluded from school during KS4 and 82% were suspended. For comparison, the rate of permanent exclusion across all individuals was 1% and the rate of suspension was 15%. Despite these extremely high

relative risks [RR 15.0 (95% CI 14.44-15.58) for permanent exclusion and 5.47 (95% CI 5.42-5.51) for suspension], the evidence that exclusion is a direct cause of violence perpetration is weak.

### **Weaknesses in the existing literature**

The majority of research demonstrating these associations is derived from American self-report longitudinal surveys, which usually lack detailed information about school performance and suffer from selection and attrition bias. In addition, their findings are vulnerable to period effects linked to changing education policy and may not be generalisable to the UK. In part, the UK evidence base is weak because a vast array of other potentially contributing factors in the lives of young people are often not taken into account. For example, a recent [DfElinke report](#) examines clustering of violent offending in school children at the local authority level but does not present any effect sizes or adjust for confounders. Family, interpersonal and community-level exposures are likely to play a significant influencing role in patterns of absence/exclusion and of violence. These influences may be mediated by, for example, [social exclusion](#) or independent of any causal impact of exclusion on violence. The correlation between exclusion and violence may also be partly explained by violence being a [cause of both exclusion and later violence](#) (this temporality [has been examined](#) recently).

These shortcomings, often the result of limitations in the data but also structural discrimination in practice and data generated processes, present a serious risk that the assertions about the link between absence/exclusion and violence are spurious or inaccurate. Understanding how these factors are related will have important policy implications as government and schools seek to identify the most effective ways to reduce youth violence. Administrative records are not a good source of information on these early potential confounders e.g. education records contain limited information on their pupils' backgrounds and police records do not contain this type of information at all. Fortunately, the rich data held in ALSPAC can address gaps that administrative data cannot.

### **1.2. Research question(s)**

Our four primary research questions are:

- (1) What is the association between persistent school absence and self-reported violent behaviour.
- (2) What is the association between persistent school absence and official sanction for serious violence.
- (3) What is the association between school exclusion and self-reported violent behaviour.

- (4) What is the association between school exclusion and official sanction for serious violence.

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

Question Number <sup>4</sup>	Interim report	Final report
1-4	Our interim report will include an overview of the data set and initial descriptive results summarising the datasets overall and our derivation of the study samples. We will also present descriptive statistics for the exposure, outcome and potential confounder measures. It will also include a theoretically informed causal diagram(s) that will underpin our research questions.	Our final report will finalise the descriptive results, and then present the final results for each of the research questions in turn. It will include a discussion of the policy implications of the results and provide new insight into the relationship between school absence and exclusion and violence.

### 1.3. Hypotheses

We do not have an *a priori* preferred hypothesis as to whether the relationships between school absence/exclusion and serious violence are causal or whether they result from confounding – this is what we will investigate in our project. Our null hypothesis is that there is no causal relationship. Through our analyses, we will determine the strength of evidence against this null hypothesis.

Based on findings from previous studies, we expect that the children in our study sample who are persistently absent and/or excluded from school in KS4 will be at increased risk of violent behaviour in late adolescence/early adulthood. However, any observed relationship between absence/exclusion and violence could be a result of confounding, which would be consistent with the null hypothesis. Confounding results from the exposure (absence or exclusion) and outcome (violence) having common cause(s). Potential confounders that we will consider include deprivation (both family and neighbourhood); poor engagement with education and

low attainment; peer group influences; substance misuse; mental illness; adverse childhood experiences.

The alternative hypothesis is that higher levels of violence in those who miss periods of school (either as a result of persistent absence or due to exclusions) is a cause of violence. The pathway from absence/exclusion from school to violence could be due to increased exposure to criminal opportunity in these children as they do not have the daily routine, structure, social rules, and protection provided by the school environment. School absence also hinders academic progress, which in turn can exclude these young people from further educational and employment opportunities – again potentially increasing their exposure to criminal opportunities and violence. If the relationship is causal, then we would expect the effect to emerge in the short- to medium-term (within 1-2 years) of persistent absences and exclusion from school.

#### 1.4. Key concepts

**Table 1.4 Definitions of key concepts**

Terms	Definition used
Self-reported violence	ALSPAC participants were asked questions relating to antisocial behaviour and crime (in the past 12 months) in questionnaires or study clinics at ages 17 and 18 years. Serious violence was defined as saying yes to: (i) Hit/kicked/punched someone else on purpose with the intention of really hurting them? or (ii) Carried a knife or other weapon for protection or in case it was needed in a fight? In addition, at age 17.5: (iii) Actually used a weapon against somebody or (iv) Used threats or actual force or violence against the other person when you stole money or property?
Official sanction for serious violence	Home Office offence codes will be used to identify police records for: (i) violence against the person, indictable only; (ii) robbery, indictable only; and (iii) possession of weapons, triable either way or indictable only. [As per Home Office definition of serious violence].
Persistent school absence	Being absent for $\geq 10\%$ of school sessions in a given academic year. [If appropriate, we will also consider a higher threshold to capture more severe absence.]
Unauthorised school absence	Any absence that the headteacher has not given permission for or where an explanation has not been provided by the parent.

Suspension	When a child is temporarily excluded from school for a set period of time. The length of time can vary but cannot exceed 45 days in one single academic year.
Permanent exclusion	When a child is permanently excluded from a school. This is the most serious sanction a school can give a child.

## 2. About the datasets

### 2.1. Overview of datasets used

This project will use data from the Avon Longitudinal Study of Parents and Children (ALSPAC), which has been linked to education data and to local police records.

ALSPAC is a birth cohort study – it has followed the same group of people from before they were born (i.e. during their mother’s pregnancy) through to the present day. ALSPAC recruited pregnant woman who had an expected due date between April 1991 and December 1992 and who lived in a defined area in and around the city of Bristol, UK. The study children (now adults in their early 30s) have been followed throughout their lives via questionnaires and clinic visits, and through record linkage.

ALSPAC has been linked to the National Pupil Database, which is a central repository of education data for children attending school in England. Data on absences and exclusions are available for ALSPAC participants during the years they were in Key Stage 4 of their education (Years 10 and 11, ages 14-16 years). Other individual-level education data available includes attainment, special educational needs, and free school meal (FSM) eligibility. School-level education data includes average measures of attainment and average FSM eligibility.

ALSPAC has also been linked to Avon and Somerset Police data, which includes records of charges, cautions and other out of court disposals for crimes committed in Avon and Somerset (a geographical area than includes the original ALSPAC recruitment area). Data are available from 2007-2021.

### 2.2. Secondary data source(s)

**Table 2.2a Dataset Description - ALSPAC**

<b>Name of dataset</b>	ALSPAC
<b>Data owner(s)</b>	University of Bristol



Type of data	Longitudinal birth cohort data
Availability of data	The ALSPAC Executive operate a <u>managed open access</u> process.
Team member(s) who will have access	Alison Teyhan, Rosie Cornish, Jasmine Rollings
Population/geographic coverage or sampling frame	ALSPAC participants born in 1991/1992 whose mothers lived in a defined area in and around the city of Bristol. Our sample will be restricted to those for whom we have permission to link to their crime and education records.
Years covered or survey waves	1991 onwards
Exclusion criteria	<p>We will exclude:</p> <ul style="list-style-type: none"> <li>- those for whom we do not have permission to link to education and police records</li> <li>- those who do not have KS4 education data</li> <li>- those who did not live in Avon and Somerset from age 16-18 years.</li> <li>- the small number who have a police record relating to serious violence prior to the start of Key Stage 4 (KS4, school years 10 and 11, age 15-16 years) and possibly those with self/parent/teacher reports of violence, depending on data availability.</li> </ul>
Expected population/sample size (following exclusion criteria) <sup>5</sup>	Based on previous work, we expect a maximum possible study sample of around 9000. The sample for the self-reported outcomes will be lower than that with the police recorded outcomes.
Documentation	<p>Cohort profile papers:</p> <p><u>Child</u></p> <p><u>Mother</u></p> <p><u>Website</u> (includes link to searchable data dictionary)</p>

---

<sup>5</sup> This may not be known at this stage of the project. Please provide your best estimate or range based on your knowledge of the dataset.

**Table 2.2b Dataset Description – *National Pupil Database***

<b>Name of dataset</b>	National Pupil Database
<b>Data owner(s)</b>	Provided by the Department for Children, Schools and Families (DCSF) (now known as Department for Education).
<b>Type of data</b>	Pupil and school level education data, including school absence and exclusion data.
<b>Availability of data</b>	The NPD data that has been linked to ALSPAC is available through ALSPAC's <u>managed open access</u> process.
<b>Team member(s) who will have access</b>	Alison Teyhan, Rosie Cornish, Jasmine Rollings
<b>Population/geographic coverage or sampling frame</b>	Children at state school in England
<b>Years covered or survey waves</b>	The ALSPAC sample span three academic years. They started KS4 (i.e. started Year 10) in 2005, 2006 and 2007 and completed KS4 (i.e. got to the end of Year 11) in 2007, 2008 and 2009 respectively.
<b>Exclusion criteria</b>	ALSPAC study children who were not at a state school in England during these years will not be in this dataset (e.g. those who are home educated, privately educated, attending a school outwith England).
<b>Expected population/sample size (following exclusion criteria)<sup>6</sup></b>	These data will be linked to the ALSPAC data; therefore, as before, we expect a study sample of around 9000 (all of whom will have KS4 records).
<b>Documentation</b>	Documentation for the education data that has been linked to ALSPAC is included in the ALSPAC Data Dictionary, which can be accessed from the ALSPAC <u>Website</u> (although note that the exclusion and absence data are not included in the dictionary). More information can be found on the <u>NPD website</u> .

**Table 2.2c Dataset Description – Avon and Somerset Police Data**

<b>Name of dataset</b>	Avon and Somerset Police Data
<b>Data owner(s)</b>	Provided by Avon and Somerset Police, extract linked to ALSPAC is owned by University of Bristol.
<b>Type of data</b>	Police recorded crime (charges, cautions, and other out-of-court disposals).
<b>Availability of data</b>	The A&SP data that has been linked to ALSPAC is available through ALSPAC's <u>managed open access</u> process.
<b>Team member(s) who will have access</b>	Alison Teyhan, Rosie Cornish, Jasmine Rollings
<b>Population/geographic coverage or sampling frame</b>	Avon and Somerset
<b>Years covered or survey waves</b>	2007-2021 (very few pre-2007 records due to paper records being used at that time).
<b>Exclusion criteria</b>	Offences committed outwith A&S are not recorded in this dataset. We will be missing some records for early, minor offences due to police rules around retention of records.
<b>Expected population/sample size (following exclusion criteria)<sup>7</sup></b>	These data will be linked to the ALSPAC data; therefore, as before, we expect a study sample of around 9000 (some of whom will have A&S Police records).
<b>Documentation</b>	A <u>data note</u> describing the linkage of ALSPAC to A&SP data has been published.

### 2.3. Primary data collection

No primary data will be collected.

### 2.4. Linking datasets

Linkage of ALSPAC to the NPD and to the A&SP data has already been achieved.

#### **ALSPAC linkage to A&SP data**

As there is no strong, persistent identifier common to both ALSPAC and the A&SP dataset, a number of personal data items available in both datasets were used to determine which

---

<sup>7</sup> This may not be known at this stage of the project. Please provide your best estimate or range based on your knowledge of the dataset.

individuals in ALSPAC had an A&SP record: forename, surname, date of birth (DoB), sex and full current and historical address(es). A combination of deterministic and probabilistic record linkage methods were used to maximise linkage coverage and minimise false matches. Full details are provided in the [data note](#).

### **ALSPAC linkage to NPD data**

In order to satisfy the confidentiality requirements of both ALSPAC and the Department for Children, Schools and Families (DCSF), the linking work was carried out by a third party (the Fischer Trust). ALSPAC supplied the Fischer Trust with names, dates of birth and current postcode for all members of the eligible cohort, whilst DCSF supplied similar details relating to all the individual datasets in the NPD. Fischer Trust then linked ALSPAC IDs to Unique Pupil Number (UPN). Checks were undertaken to identify incorrect matches. More details can be found in the ALSPAC data dictionary, available on the ALSPAC Website.

## **2.5. Access and data protection**

The data for this project will be provided by an ALSPAC data manager and will include ALSPAC clinic and questionnaire variables, education variables, and crime variables. These will be combined into one dataset either by the data managers or by the researcher, depending on what we request. Therefore, the rest of this section refers only to this one dataset and not the three datasets listed in sections 2.1 and 2.2.

### **1. Data Protection**

ALSPAC adhere to the principles of the ONS 'Five Safes framework' (safe data, safe projects, safe people, safe settings, and safe outputs).

#### ***Safe Data***

In common with all projects using ALSPAC data:

- (1) the dataset will be minimised to only include variables necessary for this project;
- (2) the data will be de-identified;
- (3) the ID variable in the dataset will be unique to this project meaning it cannot be linked to any other ALSPAC data;
- (4) potentially disclosive variables (e.g. small cell counts, precise dates) are never released to researchers.

### ***Safe projects***

A proposal for this project has been submitted to the ALSPAC Executive (October 2023) for approval. Once approved, it will be listed on the study's webpage for approved projects and assigned as ID number (known as a 'B number').

### ***Safe People***

All members of the research team are researchers experienced in working with sensitive, individual-level data. The data for this project will only be accessible by three members of the research team (AT, RC and JR). AT and RC are ONS accredited researchers, and JR will undertake this training as soon as possible after starting work on this project. AT, RC and JR have a DBS certificate dated within the last 12 months.

### ***Safe setting***

Due to the sensitive nature of the linked education and crime data, the dataset will be accessed via UKSeRP, a secure and controlled online data sharing platform hosted by Swansea University. Only three members of the research team (AT, RC, JR) will have access to the data. Their access to the data will cease at the end of this project.

### ***Safe Outputs***

All outputs will be disclosure checked by an ALSPAC Data Linkage Manager prior to release from UKSeRP.

## **2. Data Processing Roles**

The Data Controller for the information directly collected by ALSPAC is the University of Bristol. ALSPAC is also the (joint) Data Controller for information about participants collected from routine administrative sources.

## **3. Legal Basis for data processing**

ALSPAC's purpose is to conduct scientific research that aims to improve the public good and improve scientific understanding. The legal basis for using participants' information, under GDPR and the Data Protection Act 2018, is:

- 1) performance of a task carried out in the public interest (Article 6(1)(e) in the GDPR); and, where sensitive personal information is involved:
- 2) scientific or historical research purposes or statistical purposes (Article 9(2)(j) in accordance with Article 89(1)).

The GDPR defines ‘sensitive personal information’ as information that reveals a person’s racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership; and the processing of genetic data or biometric data for the purpose of uniquely identifying a person; data concerning health or data concerning sex life or sexual orientation.

This legal basis within GDPR and the Data Protection Act 2018 is separate to, and in addition to, ALSPAC seeking consent to take part in the research process, which they use to help ensure that research is ethical and complies with other applicable laws.

#### 4. Data access process

Access to ALSPAC data is via a Managed Open Access process. Researchers submit their research proposal to the ALSPAC Executive. Once the proposal has been approved, the researchers compile a detailed list of the variables they require. ‘ALSPAC Direct Users’, which AT and RC are, can then build their own datasets of questionnaire and clinic variables. This dataset will then be passed to an ALSPAC data linkage manager who will send it to UKSeRP to be uploaded to the specific folder for this project. Within UKSeRP, they will prepare the education and police data and place them in the study folder. The ALSPAC data linkage manager will perform an ID swap on all the datasets for this project so that the individual ID is unique to this project (meaning the data cannot be linked with any other ALSPAC data).

### 3. About the data

#### 3.1. List of variables

**Table 3.1: Variable definitions**

Variable abbreviation	Variable definition	Variable source	Derivation or specification
<b>Outcomes</b>			
Police recorded serious violence	Any serious violent crime police record in the 24 month period from the end of Year 10 (ages 15-17 year approx (no; yes).	A&S Police	This will be derived based on Home Office groupings and age of offence.
Self-reported serious violence	Any self-reported violence at age 17-18 (no; yes).	ALSPAC Questionnaires	Derived from questions asking: (i)

			Hit/kicked/punched someone else on purpose with the intention of really hurting them? or (ii) Carried a knife or other weapon for protection or in case it was needed in a fight? In addition, at age 17.5: (iii) Actually used a weapon against somebody or (iv) Used threats or actual force or violence against the other person when you stole money or property?
<b>Exposures</b>			
Persistent absence	Persistent absence <i>for any reason</i> during Year 10 or Year 11 (absent for $\geq 10\%$ of sessions) (no; yes). We will consider Years 10 and 11 separately.	NPD absence data	Binary variable derived from continuous measure of percentage of sessions absent.
Persistent unauthorised absence	Persistent <i>unauthorised</i> absence during Year 10 or Year 11 (absent for unauthorised reason $\geq 10\%$ of sessions) (no; yes). We will consider	NPD absence data	Binary variable derived from continuous measure of percentage of sessions absent for unauthorised reason.

	Years 10 and 11 separately.		
Suspensions	Any suspensions (fixed-term school exclusions) during Year 10 or Year 11 (no, yes). We will consider Years 10 and 11 separately. We will consider the reason for suspension if numbers allow. [Depending on data, we may also be able to consider number of exclusions/total number of days excluded for].	NPD exclusion data	Binary variable derived from exclusion data.
Permanent exclusion	Any permanent exclusion during Year 10 or Year 11 (no, yes). We will consider Years 10 and 11 separately.	NPD exclusion data	Binary variable derived from exclusion data.  [NB we expect the number of permanent exclusions to be too small for this exposure to be examined individually. We will combine with suspensions if appropriate.]
<b>Possible Confounders</b>			
<b>CHILD MEASURES</b>			
Sex	Child sex (male; female)	ALSPAC questionnaires	Standard ALSPAC measure



Age	Age at outcome time points (in months)	ALSPAC questionnaires (self-reported violence data); police data (police-recorded violence data)	Standard ALSPAC /police measures
Ethnicity	Ethnic group (White, non-White)	ALSPAC questionnaires	Standard ALSPAC measure
Early offending/violence	Any early (pre-KS4) violence	ALSPAC child, mother and teacher questionnaires	Derived from several questionnaire sources
Mental health	Depending on data, we hope to create a binary variable to identify those with mental health difficulties in early adolescence.	ALSPAC questionnaires	Binary variable derived from several ALSPAC measures.
Child substance misuse	Use of alcohol, tobacco and illicit drugs up to KS4.	ALSPAC questionnaire data.	Variables will be derived from several ALSPAC measures.
School enjoyment/attachment	Teacher they can trust, friends at school etc.	ALSPAC questionnaire data.	Variables will be derived from several ALSPAC measures.
SEN	Any special educational needs during KS4 (yes; no). [We are using this as a proxy of SEN during KS3 as SEN data are not available for those years.]	NPD data	Dichotomised from SEN variables in KS4.

School mobility	Moved school during KS3 (yes; no).	NPD data	Variable available in KS4 dataset.
ACEs	Adverse childhood experiences up to 11 years including abuse of child; abuse of mother; parental criminality; maternal mental illness; parental separation/divorce.	ALSPAC questionnaire data.	Measures will be considered individually and as a sum score.
Attainment	Educational attainment at KS2 and KS3. Probably quartiles of capped point scores.	NPD data	Variables in NPD datasets.
Commute to school	Time commute takes and method used during KS3.	ALSPAC questionnaire data.	Derived from standard measures
Peer influences	Friendships and activities with those friends.	ALSPAC clinic data at 13years.	We aim to derive a variable to define positive from negative friendships.
<b>FAMILY MEASURES</b>			
Early life SEP	We will consider various measures of early-life socio-economic position (e.g. maternal education, housing tenure, overcrowding, maternal smoking in pregnancy, family structure, financial difficulties).	ALSPAC Mother Questionnaires.	These are standard ALSPAC measures, detailed in the data dictionary.

SEP during adolescence	IDACI score (measures of neighbourhood deprivation) (probably quintiles) and FSM eligibility (no, yes) during KS3.	NPD data	Variables in NPD datasets.
Neighbourhood opinion	Maternal opinion of residential neighbourhood (safe place to live, good reputation etc). Quartiles of sum score.	ALSPAC Mother Questionnaires.	These are standard ALSPAC measures, detailed in the data dictionary.
Educational engagement	How engaged mother is with child's education	Mother and teacher questionnaires.	These are standard ALSPAC measures, detailed in the data dictionary.
<b>SCHOOL-LEVEL MEASURES</b>			
School environment	At school level: KS4 attainment, absence rates, and FSM eligibility; school size. [We are using KS4 measures here as a proxy for KS3 due to data availability].	NPD data	Quartiles of each measure (e.g. for attainment Q1=low attainment, Q4=high attainment).
<b>NEIGHBOURHOOD-LEVEL MEASURES</b>			
Residential neighbourhood environment	Deprivation and crime quintiles.	IMD and crime domain from neighbourhood statistics. IDACI from NPD.	Quintiles will be derived by linkage data managers.

School neighbourhood environment	Deprivation and crime quintiles.	IMD and crime domain from neighbourhood statistics.	Quintiles will be derived by linkage data managers.
----------------------------------	----------------------------------	---	---

### 3.2. Measurement of key concepts

**Table 3.2 Measurement of key concepts**

Concept <sup>8</sup>	How the concept will be measured and encoded
Self-reported violence	This has been measured in ALSPAC questionnaire and clinic data at ages 17 and 18. Serious violence will be defined as saying yes to: (i) Hit/kicked/punched someone else on purpose with the intention of really hurting them? or (ii) Carried a knife or other weapon for protection or in case it was needed in a fight? In addition, at age 17.5: (iii) Actually used a weapon against somebody or (iv) Used threats or actual force or violence against the other person when you stole money or property?
Official sanction for serious violence	The A&SP data linked to ALSPAC includes Home Office offence codes for each offence. From this, we will identify which records at ages 15-17 years relate to a serious violent crime: (i) violence against the person, indictable only; (ii) robbery, indictable only; and (iii) possession of weapons, triable either way or indictable only. [As per Home Office definition of serious violence].
Official sanction for any offending	If numbers of police records for serious violence are too low, we will identify records at 15-17 years for any offence.
Persistent school absence	NPD absence data for KS4, provided by the DfE, has been linked to ALSPAC. Persistent absence will be defined as being absent for $\geq 10\%$ of school sessions in an academic year.
Unauthorised school absence	NPD absence data for KS4 specified if an absence was authorised or unauthorised.
Suspension	NPD fixed-term exclusion data for KS4 has been linked to ALSPAC.
Permanent exclusion	NPD permanent exclusion data for KS4 has been linked to ALSPAC.

### 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**

We do not have absence and exclusion data before KS4, meaning we will not be able to examine trends in school attendance over time.

At KS4, we know the percentage of sessions that a pupil was absent in an academic year but not the date(s) that they were absent. This means we will not be able to consider if any violence coincided with or immediately followed a period of absence.

We have very few police records prior to 2007 due to paper records being used before this date, and the fact that the police can only retain records if they have justification to do so (as per MoPI rules).

At our project planning focus group, the education experts raised several factors that they felt could be important to consider in the relationship between absence/exclusion and violence that we do not have in our datasets. These include measures of child exploitation, whether the child is a 'school refuser', whether they have experienced discrimination, measures of school ethos, and local neighbourhood characteristics such as things to do outside of school and local employment opportunities.

When we conduct our preliminary analyses, it may become clear that a small number of the variables we are considering as potential confounders do not have adequate numbers/variability in our sample and so will not be able to be included in our main analyses.

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

As ALSPAC is a longitudinal study, there is attrition and it is known that those from more disadvantaged backgrounds are more likely to drop-out of the study over time. From previous work (not yet published) we know that very few of those with a police record for violence have ALSPAC questionnaire or clinic measures during adolescence. We will include descriptions of our study samples and missing data in both our interim and final reports.

#### **(iii) Potential approaches of mitigating the impact of missing data**

We will consider how best to deal with missing data once we have accessed the data and will consider the merits of both complete case analyses and multiple imputation. We will explain what method we are planning to use in our interim report.

### 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 police data. Details are given in the [Data Note](#). 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. Examples of this include the disproportionate use of Stop and Search on Black, Asian and Minority Ethnic communities and variations in the rate of reporting of crime across communities and demographic groups. Bias may also be introduced through the data linkage process if participants with a criminal record are, in general, less active in ALSPAC, resulting in their identifier information (e.g. current name and address) held by the study being out of date.

Responses to the questionnaire data may be affected by social desirability bias or recall bias.

Those for whom ALSPAC has permission to link to their education and police records may differ from those who have opted out of linkage or who have not received a consent pack. However, opt out rates for crime and education linkage are very low (<4%).

With regards the YEF's focus on race equity, it is important to note that ALSPAC is predominately a White UK cohort (>95%), which largely reflects the demographics of the recruitment area at the time the study began in the early 1990s. Therefore, our data are not suitable for examining ethnic differences in the relationship between absence/exclusion and violence.

## 4. About the analysis

### 4.1. Overview of analytical approach

Our descriptive analyses will include an overview of the dataset, the derivation of our specific study samples for each research question, a summary of each of our exposures and outcomes, and an examination of potential confounder variables (whether they are associated with both exposure and outcome, and whether numbers are adequate).

Then to answer our four main research questions we will use multilevel regression models, which will account for the clustering of our sample within schools. We will consider the unadjusted association between the exposures (persistent absence/exclusion) and outcomes (self-reported violence/official sanction for violence) and then examine the impact of adjusting for both early life and adolescent factors at the individual, family, school and neighbourhood level.

We have decided to focus on absence and exclusion during Year 10 rather than the whole of KS4. This is because for any serious violence that takes place during KS4, we would not know if it preceded, coincided, or followed periods of absence or exclusion. By considering Year 10 as our exposure period, we can include violence outcomes from the end of Year 10 onwards, knowing that they definitely *followed* any absence/exclusion in Year 10. We are going to initially consider a 24 month period for police recorded violence: we know that the police do not detect or record every occurrence of violence. Perpetrators are more likely to be detected by police if their violence is frequent or serious. Therefore we need a long enough outcome period for there to have been a chance for the police to detect the violence, while balancing that with the fact that if exposure/absence are causally related to violence then we would expect a short time frame from exposure to outcome. We will make a final decision on our outcome period once we have explored our data.

With the suspensions/exclusions, we will consider number of suspensions during KS4 as an exposure if the data allows as we hypothesise that if the relationship is causal then there should be a dose response relationship (i.e. greater risk of violence in those with more exclusions/suspensions).

#### 4.2. Approach to addressing research question(s)

**Table 4.1 Research questions: approach and methods**

<b><i>Research question</i></b>	<p>(1) What is the association between persistent school absence and self-reported violent behaviour.</p> <p>(2) What is the association between persistent school absence and official sanction for serious violence.</p> <p>(3) What is the association between school exclusion and self-reported violent behaviour.</p> <p>(4) What is the association between school exclusion and official sanction for serious violence.</p> <p>[Questions 1-4 are all shown in this table as we will take the same general approach to all of them. Any differences are specified].</p>
<b><i>Hypothesis, if relevant</i></b>	<p>Null hypothesis: there is no causal association between school absence/exclusion and self-reported violent behaviour/police records for violence.</p>

<p><i>What will you be able to say by the interim report</i></p>	<p>For each research question, we will be able to state the provisional sample size (with flow chart showing derivation of study sample), and highlight how this sample compares to the overall ALSPAC sample in terms of key demographic and confounder variables. We will be able to summarise the prevalence of persistent absence (both all absence and unauthorised) and self-reported violent behaviour in the sample. We will identify potential confounders in the relationships between each exposure and outcome.</p>
<p><i>Descriptive analysis, if relevant</i></p>	<p>For these research questions, In the interim report we will include summary statistics detailing:</p> <ol style="list-style-type: none"> <li>(1) The provisional size of the study sample for this specific question.</li> <li>(2) The main demographic variables for this study sample.</li> <li>(3) The prevalence of self-reported violence.</li> <li>(4) The prevalence of persistent school absence and persistent unauthorised school absence measures.</li> <li>(5) Which variables are associated with both the absence and violence measures (and therefore are potential confounders).</li> <li>(6) The extent of missing data.</li> </ol>
<p><i>Models, specifications and statistical techniques used, if relevant</i></p>	<p>For research question 1 and 3 we plan to use multilevel (level 1 – individual, level 2 – school) logistic regression models to examine the relationship between persistent school absence/exclusion in Year 10 and any self-reported violence between the ages of 16 and 18 years.</p> <p>For research question 2 and 4 we plan to use multilevel Poisson regression to examine the relationship between persistent school absence/exclusion in Year 10 and any police-recorded violence in the 24 month period from the end of Year 10 (approximately ages 15-17 years).</p>



	<p>[Depending on data, we may also consider persistent absence/exclusion during Year 11 as a separate exposure.]</p> <p>Multilevel models will be used as we know our sample are clustered in secondary schools. However, if the variance in self-reported violence between schools is negligible then we will consider the use of single-level models (with the exception of models with school-level covariates, where multilevel models will always be used).</p> <p>Our modelling strategy will be refined while we conduct our initial descriptive analyses, but our provisional plan is:</p> <p>Model 1 – unadjusted</p> <p>Model 2 – adjusted for sex and ethnicity</p> <p>Model 3 – Model 2 + early life SEP + prior offending</p> <p>We will then build on Model 3 to examine (separately):</p> <p>Model 4 - Child mental health and substance use</p> <p>Model 5 – School attainment and engagement (both child and parent)</p> <p>Model 6 – Peer influences/friendships</p> <p>Model 7 – School environment</p> <p>Model 8 – neighbourhood environment</p> <p>(NB models 4 to 8 will likely each involve several models, e.g, 4a, 4b, 4c as there are several covariates to examine in each of these broad areas. Exact variable choice will be decided after our initial examination of the data).</p> <p>Model 9 – fully adjusted, including key variables from models 4-8.</p>
<p><i>Estimating equation, if relevant</i></p>	<p>The included variables will depend on what measures we identify are potential confounders in our descriptive work.</p>

<b><i>What does the approach need to succeed (constraints/assumptions)?</i></b>	Variables will only be able to be included if there is adequate variability/numbers in each category. Variables not meeting this requirement will be excluded.
<b><i>Uncertainty and inference</i></b>	Results of the logistic regression models will be given as unadjusted and adjusted odds ratios with 95% confidence intervals and corresponding p-values.
<b><i>Robustness checks</i></b>	We will examine model-fit statistics.
<b><i>Subgroup you intend to study</i></b>	We do not intend to undertake subgroup analyses. But will test for interactions with sex if numbers allow.
<b><i>Changes to the analysis</i></b>	As stated, variables will only be included in models where numbers are adequate. This should minimise convergence issues but where these arise, the researcher (JR) will consult with the statistician on this project (RC). Similarly, RC will be consulted on any other changes to the planned analyses.

## 5. Project management

### 5.1. Risks and mitigations

**Table 5.1 Risks and mitigations**

<b>Number</b>	<b>Risk</b>	<b>Likelihood (Low/Medium/ High)</b>	<b>Mitigation</b>
1	Fail to get approval from ALSPAC Exec for this project	Low	This project has already been discussed with the Data Linkage Manager and approved in principle. If Exec have any concerns (unlikely), we will revise our application to address them.

2	UKSeRP access issues	Very low for long-term issues; medium for one off, short-term (<1 day) issues.	If they occur, these will be reported to UKSeRP and the ALSPAC data linkage team as soon as possible so that can be resolved. JR will be advised to do any parts of her work that don't require to be in UKSeRP outwith that environment so that she always has something to work on.
3	Number of participants with both exclusion/absence records and self-reported violence or police violence at ages 15-17 small	Medium	We have chosen to focus our outcome time period for police-recorded violence on the 24 month period immediately following the end of Year 10. If small numbers with police-recorded violence mean we are lacking statistical power, we will consider the merits of (1) increasing the length of this outcome time period and (2) including all offending in addition to serious violent offending. Similarly, for the self-reported violence, in the event of small numbers, we will consider inclusion of measures available at older ages.

## 5.2. Timeline

**Table 5.2 Timeline**

Date	Activity	Staff responsible/leading
Oct/Nov 2024	Preparation of data set	AT (and ALSPAC data linkage managers)

Nov-Jan	Cleaning of data, descriptive analyses.	JR (with AT, RC and IB advising)
February 2024	Interim report to YEF	JR (with AT, RC and IB also contributing).
March-June 2024	Analytical models	JR (with AT, RC and IB advising).
July 2023	Final report to YEF	JR and AT (with RC and IB also contributing).

## 6. References

Links have been provided within document.