Project title: Examining adverse and positive childhood experiences, and neighbourhood violent crime, as determinants of youth violence: Analyses of data from the Millennium Cohort Study

Secondary Data Project Analysis Plan

Evaluating institution: University College London, Centre for Longitudinal Studies

Principal investigator(s): Dr Aase Villadsen

Analysis Plan for YEF Secondary Data Analysis Projects

Project summary

Project title	Examining adverse and positive childhood experiences, and neighbourhood violent crime, as determinants of youth violence: Analyses of data from the Millennium Cohort Study
Research Team	Dr Aase Villadsen; Dr Nicolas Libuy; Prof Emla Fitzsimons (University College London, Centre for Longitudinal Studies)
Principal investigator	Dr Aase Villadsen
Analysis plan author(s)	Dr Aase Villadsen; Dr Nicolas Libuy; Prof Emla Fitzsimons
Overarching research question	How do adverse and positive childhood experiences, and violent crime in the neighbourhood, contribute to youth violence?
Supporting research question(s)	1) How do positive and negative childhood experiences relate to violent crime in adolescence? 2) Is neighbourhood crime an important determinant of youth violence? 3) How do neighbourhood crime and childhood experiences (good and bad) interact?
Dataset(s) to be used	Millennium Cohort Study; Police recorded crime

Population characteristics	Young people aged 17 born around the Millennium (Sep 2000-Jan 2002).
Years data spans	2001 to 2018
Geographic coverage	England and Wales
Primary outcome(s) investigated	Youth violence as measured through self-reports
Main method(s) to be used or tested	Multivariate logistic regression

About this document

This document should be completed at the start of all YEF secondary data analysis projects. It should be written for a technical analysist audience, who have no prior knowledge of the research being conducted. Its aim is to increase transparency, minimise bias and ensure continuity if there are any changes in the research team. The analysis plan will be reviewed by the YEF and we may decide to have it peer reviewed. It may also be published on the YEF website.

All secondary data analysis projects will have an interim and final reporting stages. The interim report will include the results from initial exploration of the dataset(s) being used and from feasibility testing of alternative methods in addressing the research question(s). The final report will include all results and analysis that address the research question(s). For projects that are at an early stage in exploring what's possible, this template should be completed with a view to understanding how feasible the alternative methodologies are, and what criteria will be used to assess what's viable.

Any guidance notes (in italics) can be deleted on completion and replaced with the actual text which should not be in italics and instead in justified black Calibri font size 12 with 8pt spacing before and after and multiple 1.15 line spacing.

Analysis plan history

Version	Date	Reason for revision
1.X [*latest*]		
1.1		
1.0 [*original*]		*[leave blank for the original version]*

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

Table of contents[1]

[1] To update the table of contents, right click on the table > select update field > select update entire table.

1. About the project

1.1. Background to the project

Youth offending statistics[1] in England and Wales overall has shown a significant decrease in the last decade; however, offences classified as 'violence against the person' remain high, with around 15,000 such offences being carried out in 2020 by young people aged 10 to 17 years. Moreover, knife and other weapon offences involving young people are nearly 50% higher in 2020 than five years previously. A better understanding of the risk and protective factors associated with these behaviours, using a contemporaneous cohort of young people, is a crucial first step in informing policy makers and practitioners on routes to prevention and intervention.

In this research we propose to focus on adverse childhood experiences (ACEs), positive childhood experiences (PCEs), and neighbourhood violent crime levels, as well as how these dimensions interact with each other in predicting youth violence. The family and home environment are undoubtedly one of the most important influences on the adjustment and development of children and young people, including the development of aggressive and violent behaviours.[2] Whilst single adverse aspects of the family environment have been identified - including poverty, maltreatment, family violence, low parental involvement and warmth, harsh parenting, parental stress and poor mental health, and lack of supervision[3] - the study of ACEs focuses on the accumulation of multiple aspects of family dysfunction.[4-8] At the other end of the spectrum, the concept of PCEs is concerned with resilience, and the accumulation of familial factors as well as factors across other domains of the child's life that can protect against detrimental outcomes. Research has indicated that PCEs benefit multiple outcomes, including physical health, mental health, school achievement, teen pregnancy, and youth offending.[9-13] Furthermore, PCEs may be able to moderate or buffer the negative effect of ACEs, with evidence from both the field of mental health and criminology.[9 14-16]

Another potential influence on youth violence is the neighbourhood children grow up in.[17] Whilst there is a multitude of dimensions to the neighbourhood, the level of violent crime seems particularly relevant to youth violence. Some research has indicated that being exposed to violence as a victim, or seeing someone else being victimised, is associated with an increase in young people's likelihood of engaging in violence and carrying weapons.[18] Although, understanding the causal role of the neighbourhood is immensely challenging as neighbourhood characteristics tend to be highly correlated with individual and family characteristics, including ACEs and PCEs (i.e. selection bias), [19] which may be the more causal drivers of outcomes such as youth violence. Others have suggested that neighbourhood risks may be associated differentially with youth offending depending on exposure to ACEs and PCEs. A US study found that neighbourhood disadvantage was associated with an increase in

serious crime offending in those young people who had a balanced mix of risk and protective factors, or scored high on protective factors; whilst young people with a high risk score engaged in serious crime offences at a similar rate regardless of the level of deprivation in their neighbourhood.[20] Although this was an unanticipated finding, study authors suggest that a possible interpretation might be that activities in adolescence increasingly take place in public settings, and this may mean that protective aspects of the family settings may be overpowered by the influence of a disadvantaged neighbourhood.

To date, few studies have applied the ACEs framework to the study of youth violence, [21-23] with even fewer have examined PCEs,[14] and existing studies are based on mostly retrospective rather than prospective data. There remains a major dearth of evidence in this area for the UK. There is also a lack of work examining young people's exposure to levels of violent crime in the neighbourhood and its association with youth violence. By linking area based crime data to the MCS, the proposed research presents an opportunity to examine the role of area violence in combination with a wealth of other influential factors in the development of youth violence, using a large nationally representative UK cohort study. It is also important to understand in a UK context how levels of violent crime in the neighbourhood may potentially interact with ACEs and PCEs in relation to youth violence.

Additionally, it is important to understand whether ACEs, PCEs and neighbourhood crime affect groups of young people equally in terms of offending, especially across major social categories such as gender and ethnicity. Studies have suggested that males are more adversely influenced by ACEs than females [24 25] and also neighbourhood effects appear more pronounced in males than females[26], whilst for PCEs there is support that this is more protective for females [15]. As for ethnicity, previous studies based on US data have generally indicated that ACEs have a stronger association with youth delinquency in ethnic minorities than in white populations.[27]

Providing answers to this study's research questions is of significant value to those tasked with reducing youth violence, as it can help inform where to direct efforts and resources - ultimately leading to a reduction in youth violence.

By identifying ACEs, the project has the promise to provide current information on some of the predictors of youth violence in England and Wales, which is a first step in informing early prevention strategies. Equally important, but far less frequently examined in previous studies, is the identification of PCEs, which highlight protective factors, and which can inform strategies of support to young people. The potential of PCEs to reduce youth violence in the face of a high level of ACEs is an important aspect to examine, with results having the potential to make a significant impact on policy, practice, and intervention. To what extent area levels of violence matter for individuals' propensity for engagement in violent crime, over and above

their family levels risks and protective factors, as well as interacting with these, is especially important for understanding whether levers of interventions and prevention should be focused on the neighbourhood or elsewhere.

1.2. Research question(s)

How do positive and negative childhood experiences relate to youth violence?

1. Which single and cumulative ACEs are associated with youth violence, and what are the magnitudes of the associations?
2. Which single and cumulative PCEs are associated with youth violence, and what are the magnitudes of the associations?
3. Do PCEs attenuate the association between ACEs and youth violence?

Is neighbourhood crime an important determinant of youth violence?

4. To what extent is neighbourhood violent crime using police.uk data a valid and reliable measure?
    a. To what extent is the measure of neighbourhood crime using the police.uk data consistent with the measure of neighbourhood crime using the Index of Multiple Deprivation (IMD)?
    b. Does the police.uk violent crime measure show the same national level trend over time (2011-2021) as the published Home Office statistics on crime in England and Wales?
5. Are rates of violent crime in one's neighbourhood associated with youth violence?

How do neighbourhood crime and childhood experiences (good and bad) interact?

6. Does the association between ACEs and youth violence differ for those in neighbourhoods with high versus low levels of violent crime?
7. Does the association between PCEs and youth violence differ for those in neighbourhoods with high versus low levels of violent crime?
8. Do PCEs attenuate the association between ACEs and youth violence more in low-crime areas or in high-crime areas?
9. Do ACEs amplify the association between neighbourhood violent crime and youth violence?
10. Do PCEs attenuate the association between neighbourhood violent crime and youth violence?

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

Question Number	Interim report	Final report
N/A	Descriptive statistics of the sample and study variables	Adjustments, edits and additions to descriptive statistics
1	Proposed ACEs measures will be described and examined in relation to youth violence. Main results will be shown in tables or graphs and briefly summarised.	Presentation in graphs or tables and write up in full of results using the agreed ACEs measures. Final methods, interpretation and discussion of results will be provided.
2	Proposed PCEs measures will be described and examined in relation to youth violence. Main results will be shown in tables or graphs and briefly summarised.	Presentation in graphs or tables and write up in full of results using the agreed PCEs measures. Final methods, interpretation and discussion of results will be provided.
3	Proposed ACEs and PCEs will be examined in terms of how PCEs may attenuate the association between ACEs and youth violence. Main results will be shown in tables or graphs and briefly summarised.	Presentation in graphs or tables and write up of results using the agreed ACEs and PCEs measures. Final methods, interpretation and discussion of results will be provided.
4	Testing and recommendation for whether police.uk data can provide a valid and reliable measure of neighbourhood violent crime for use in the current study.	The agreed data source and variables from the interim stage will be used in analyses. All analyses presented in graphs and tables and written up in full. Final methods, interpretation and discussion of results will be provided.
5-10	No results will be included in the interim report.	All analyses presented in graphs and tables and written up in full. Final methods, interpretation and discussion of results will be provided.

1.3. Hypotheses

The background (Section 1.1) sets out the literature and justification for the project and our hypotheses are built around this. In summary, ACEs and neighbourhood levels of crime are considered risk factors for children's development and adjustment, and therefore in this

research project we hypothesise higher levels of these be associated with a higher prevalence of youth offending. This is opposed to PCEs which are considered protective factors, as these are thought to support positive development and adjustment in children and young people, and we, therefore, hypothesise that higher levels of PCEs are associated with a lower prevalence of youth violence.

- Q1: We expect single and cumulative ACEs to be associated with a higher likelihood of youth violence.
- Q2: We expect single and cumulative PCEs to be associated with a lower likelihood of youth violence.
- Q3: We hypothesise that a higher level of cumulative PCEs will attenuate the association between cumulative ACEs and youth violence.
- Q4: a) We expect LSOAs to have a similar ranking in terms of level of neighbourhood crime measured using police.uk data as in the IMD data. b) We expect the police.uk measure of neighbourhood crime to show the same trend over time as the published Home Office national figures for crime.
- Q5: We hypothesise that a higher level of violent crime in the neighbourhood is associated with a higher likelihood of youth violence.
- Q6: As a high level of violence in the neighbourhood is a potential additional risk factor, we hypothesise that ACEs have a stronger association with youth violence in neighbourhoods with high levels of violent crime compared to areas with lower levels.
- Q7: We hypothesise that the association between PCEs and youth violence will be weaker in neighbourhoods with a high level of violent crime. This is because neighbourhood violent crime is a potential risk factor and may undermine the protectiveness of PCEs for youth violence.
- Q8: Because neighbourhood violent crime is a potential risk factor, the extent to which PCEs reduce the association between ACEs and youth violence is hypothesised to be lower in areas with high levels of violent crime.
- Q9: Because ACEs are likely to be an additional risk factor, we hypothesise that neighbourhood violent crime will have a stronger association with youth violence for those with a high level of ACEs.
- Q10: Because PCEs are likely a protective factor, we hypothesise that neighbourhood violent crime will have a weaker association with youth violence for those with a high level of PCEs.

1.4. Key concepts

Table 1.4 Definitions of key concepts

Terms	Definition used
Adverse childhood experiences (ACEs)	These are experiences in childhood which are thought to be risk factors for adverse outcomes. The exact factors will be explored and developed in the initial phase of the project, and we plan to draw heavily on previous research in the field including studies that use the MCS. Our proposed AECs measures will be presented in the interim report.
Positive childhood experiences (PCEs)	These are the experiences in childhood which are thought to protect against adverse outcomes. In our definition and measurement of PCEs we will draw on previous literature. Our proposed PCEs measures will be presented in the interim report.
Neighbourhood violent crime	Subject to reliability and validity checks (Q4), we will use police.uk provided data and focus on violent crimes. The total number of violent crimes will be summed up for each Lower Layer Super Output Area (LSOA) which will define the neighbourhood, and this will then be divided by the estimated population for the LSOA. Our proposed measure of neighbourhood violent crime will be presented in the interim report.
Youth violence	Three aspects of youth violence will be examined as separate outcomes [28]. Information will be combined across age 14 and 17 for each of the outcomes. 1) *Assault* is defined as pushing, shoving, hitting, slapping or punching someone. 2) *Weapons involvement* is defined as carrying or using a weapon. 3) *Gang involvement* is defined as current or past membership of a street gang.

2. About the datasets

2.1. Overview of datasets used

Data from the UK Millennium Cohort Study (MCS) will be our primary dataset. This includes a large sample of around 19,000 individuals born in the UK around the Millennium who are aged around 17 years at most recent follow-up. We will use data from all sweeps from age 9 months to age 17 years (2000-2018). We restrict the sample to those living in England and Wales.

In addition, we will make use of police provided crime data at neighbourhood level which will be linked to the MCS. We will link police recorded crime data obtained from police.uk (2011-2013), which covers 43 police forces in England and Wales.

Furthermore, the Index of Multiple Deprivation in England (2010) and in Wales (2011) will be used to check the consistency of the police.uk data (2011-2013). This will mean a linking between police.uk data and IMD data at the LSOA level. Both datasets will then be linked to the MCS in order to examine the consistency between datasets.

Published open source Home Office statistics on police recorded violent crime at police force level will also be used to run additional checks on the police.uk crime measure. This data will however not be linked to the MCS or any other data source.

Finally, data on ONS LSOAs population estimates will be used and linked in order to derive crime rates for each LSOA.

2.2. Secondary data source(s)

Table 2.2a Dataset Description – Millennium Cohort Study

Name of dataset	Millennium Cohort Study
Data owner(s)	UCL Centre for Longitudinal Studies
Type of data	Longitudinal birth cohort study
Availability of data	Freely available to researchers via the UK Data Service under end user licence.
Team member(s) who will have access	Aase Villadsen; Nicolas Libuy
Population/geographic coverage or sampling frame	Individuals born in the UK in 2000-2002 sampled through the near universal child benefit register.
Years covered or survey waves	2001-2018 (seven waves in total), at age 9 months, age 3, 5, 7, 11, 14 and 17 years.
Exclusion criteria	Sample includes those living in England and Wales at the time last observed, thereby excluding those in Scotland and Northern Ireland.
Expected population/sample size	Around 10,000 young people will be included in the final analytical sample

(following exclusion criteria)	
Documentation	**Sweep 1-5 (9 month to age 11):** https://cls.ucl.ac.uk/wp-content/uploads/2017/07/MCS-Guide-to-the-Datasets-022014.pdf **Sweep 6 (age 14):** https://cls.ucl.ac.uk/wp-content/uploads/2018/10/mcs6_user_guide_28march2017.pdf **Sweep 7 (age 17):** https://cls.ucl.ac.uk/wp-content/uploads/2022/05/MCS7-user-guide-Age-17-ed2.pdf

Table 2.2b Dataset Description – Police Recorded Crime (police.uk)

Name of dataset	**Police Recorded Crime in England and Wales**
Data owner(s)	police.uk
Type of data	Administrative data of crimes reported to the police for each of the 43 territorial police force in England and Wales, plus the British Transport Police.
Availability of data	Data at police force level is freely available under the Open Government Licence v3.0 and can be downloaded from the police.uk website. It is available by police force and LSOA level for year 2011 onwards. Offences are categorised into 11 categories. These data have previously been linked to the MCS for another project, and linked data is expected to be deposited at UKDS in 2023. To link these data to the MCS, along with the HO provided data, a new application will be made to the Centre for Longitudinal Studies Data Access Committee (CLS DAC). See section 2.5 below for details on this.
Team member(s) who will have access	Aase Villadsen; Nicolas Libuy

Population/geographic coverage or sampling frame	All crimes reported to the police in England and Wales by neighbourhood/LSOA level.
Years covered or survey waves	We require data covering years 2011-2021.
Exclusion criteria	None
Expected population/sample size (following exclusion criteria)	All 43 police force areas in England and Wales will be included, representing all 33,755 LSOAs in these two UK countries.
Documentation	Information on this data and documentation: https://data.police.uk/about/

Table 2.2c Dataset Description – Index of Multiple Deprivation (IMD)

Name of dataset	**Index of Multiple Deprivation (IMD) in England and Wales**
Data owner(s)	Gov.uk, Stats Wales
Type of data	National statistics
Availability of data	These national statistics data are freely available and can be downloaded straightforwardly from the websites linked below in the documentation section.
Team member(s) who will have access	Aase Villadsen; Nicolas Libuy
Population/geographic coverage or sampling frame	These data contain all LSOA areas in England and Wales. Each LSOA has an overall ranking across the total Index of Multiple Deprivation, and a ranking for each domain of the index.
Years covered or survey waves	IMDs are constructed with some years interval. We require the English 2010 version and the Welsh 2011 version as

	these are most consistent with the police.uk data covering years 2011-2013, and with which we want to compare.
Exclusion criteria	None
Expected population/sample size (following exclusion criteria)	All 33,755 LSOAs across England and Wales.
Documentation	Information on this data and documentation England: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010 Wales: https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/Archive/WIMD-2011

Table 2.2d Dataset Description – Police Recorded Crime (Home Office)

Name of dataset	**Police Recorded Crime in England and Wales**
Data owner(s)	Home Office
Type of data	Administrative data of crimes reported to the police for most of the 43 territorial police force in England and Wales, plus the British Transport Police.
Availability of data	Data at police force level is freely available as open data and national figures of different types of crime can be calculated straightforwardly for each year.
Team member(s) who will have access	Aase Villadsen; Nicolas Libuy
Population/geographic coverage or sampling frame	All crimes reported to the police in England and Wales by neighbourhood/LSOA level, with exception of some police forces who do not submit their figures, including Greater

	Manchester Police, West Midlands Police and a number of other forces.
Years covered or survey waves	We require data covering years 2011-2021 as this is the period over which we wish to compare trends in violent crime with police.uk data.
Exclusion criteria	None
Expected population/sample size (following exclusion criteria)	Most police force areas in England and Wales will be included, but with exclusions as noted above.
Documentation	Data user guide: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/560132/pprc-user-guide-oct16.pdf General website: https://www.gov.uk/government/statistics/police-recorded-crime-open-data-tables#full-publication-update-history

2.3. Primary data collection

No primary data will be collected.

2.4. Linking datasets

Linking police recorded crime data from police.uk to IMD data

We will link IMD data and police.uk data at the LSOA-level. To ensure accurate comparisons, we will match the periods used to create both datasets as closely as possible, even though they may differ.

Linking police recorded crime data from police.uk and IMD data to MCS

LSOA-level data from police.uk linked with IMD data will then be linked to the MCS using the LSOA of MCS participants' residences (subject to DAC approval). The LSOAs where participants live at each MCS sweep is derived by the Centre for Longitudinal Studies using the residential postcode of MCS participants. This information will be accessed through an application to the Centre for Longitudinal Studies Data Access Committee. The LSOA

identifiers in MCS are available for all productive surveys, which will allow us to link police.uk data for 100% of MCS participants in England and Wales.

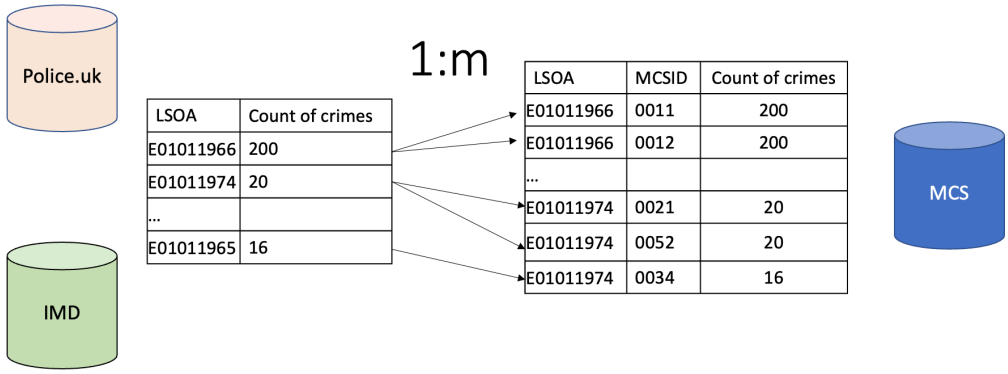
Linking ONS LSOAs population estimates to MCS

We will link ONS annual LSOA population estimates to MCS data using the LSOA of MCS participants' residences. Population estimates at LSOA level will be used as the denominator of crime rates. Initially we will link population estimates for year 2012 as it corresponds to the survey fieldwork period of MCS sweep 5 (age 11).[29]

Linkage Method

As shown in the diagram below, the linkage between police data and MCS will be 1 to many (1:m) because IMD data and police.uk is aggerated at LSOA-level using a unique LSOA identifier and there are multiple MCS participants per LSOA.

Diagram indicating the linkage between police data and MCS



2.5. Access and data protection

Accessing MCS Data

Data from the MCS are freely available to the research community through the UK Data Service.[30] Researchers who will be working on this project already have full access to all available data from the initial survey at 9 months, and at each follow-ups at ages 3, 5, 7, 11, 14 and 17. In addition, project researchers have extensive experience in managing and analysing MCS data, including work on offending behaviours at age 17.[28]

Accessing police-recorded crime data from police.uk

These data are freely at neighbourhood/LSOA level and can be straightforwardly downloaded from the date.police.uk website under the Open Government Licence v3.0.

Accessing IMD data

Publicly available IMD data for England (2010) and Wales (2011) will be downloaded from the Official Government website.

Data linkage

For the linkage between MCS data and police.uk recorded area crime data, approval for the proposed study will be sought from the Centre for Longitudinal Studies Data Access Committee (https://cls.ucl.ac.uk/data-access-training/data-enhancements/). An application will be made to access the postcodes of MCS participants. One of the project researchers has previously gained approval for a similar data linkage for carrying out some preliminary work in the area.

Data protection

All researchers at UCL undergo training in handing and processing of confidential data by annually completing the NHS Digital's Data Security Awareness (NHSD) course. Although all data in the MCS remain fully anonymised, once data have been linked to additional sources, they are deemed especially sensitive. All analyses will therefore be carried out via the UCL Data Safe Haven (DSH), which conforms to the NHS Digital's Information Governance Toolkit. Data will be retained in DSH for a period of 10 years after publication of results in line with UCL research data policy, after which it will be securely deleted within DSH using the Cipher Security Tool.

As researchers, we will adhere to the end user agreement of the UK Data Service https://ukdataservice.ac.uk/app/uploads/cd137-enduserlicence.pdf. This involves a range of conditions; one of them is to preserve the confidentiality of, and not attempt to identify, individuals, households or organisations in the data. Due to the large sample size of the MCS, it is unlikely that any analyses will produce small cell numbers; however, we will adopt the rule of thumb to not report any results with table cell counts of 10 or less.

The processing and analysis of these data is fully within the UK GDPR framework on the basis that MCS participants have given their consent for their data to be used for research purposes, including linkage with other data. An additional legal basis is that research is a public task, and the data processing is necessary for us to perform this task. This includes the analyses of special categories of data which in this project includes gender and ethnicity. See the following link for the legal basis of data processing:

https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/#what

3. About the data

3.1. List of variables

Table 3.1: Variable definitions

Variable abbreviation	Variable definition	Variable source	Derivation or specification
assault14	**Assault age 14.** A binary variable that indicates if the individual has assaulted someone in the past year at age 14. (0=no, 1=yes)	MCS	From a single item in age 14 questionnaire.
assault17	**Assault age 17.** A binary variable that indicates if the individual has assaulted someone in the past year at age 17. (0=no, 1=yes)	MCS	From a single item in age 17 questionnaire.
assault1417	**Assault age 14/17.** A binary variable that indicates if the individual has assaulted someone in the past year, either at age 14 or at age 17. (0=no, 1=yes)	MCS	Derived from variables: assault14 assault17
used14	**Weapon use age 14.** A binary variable that indicates if the individual at age 14 has used a weapon in the past year (0=no, 1=yes)	MCS	From a single item in questionnaire.
carried14	**Weapon carrying age 14.** A binary variable that indicates if the individual	MCS	From a single item in age 14 questionnaire.

	at age 14 has carried a weapon ever (0=no, 1=yes)		
weapon14	**Weapons involvement age 14.** A binary variable that indicates if the individual at age 14 has carried a weapon ever or used a weapon in the past year (0=no, 1=yes)	MCS	Derived from variables: used14 and carried14
used17	**Weapon use age 17.** A binary variable that indicates if the individual at age 17 has used a weapon in the past year (0=no, 1=yes)	MCS	From a single item in ag 17 questionnaire.
carried17	**Weapon carrying age 17.** A binary variable that indicates if the individual at age 17 has carried a weapon in the past year (0=no, 1=yes)	MCS	From a single item in age 17 questionnaire.
weapon17	**Weapons involvement age 17.** A binary variable that indicates if the individual at age 17 has carried or used a weapon in the past year (0=no, 1=yes)	MCS	Derived from variables: used17 and carried17
weapon1417	**Weapons involvement age 14/17.** A binary variable that indicates is if the individual has: a) carried a weapon ever at age 14 or in the past year at age 17. OR b) used a weapon in the last 12 months at age 14 or 17. (0=no, 1=yes)	MCS	Derived from variables: weapon14 and weapon17

gang14	**Gang involvement age 14.** A binary variable that indicates if the individual currently at age 14 or in the past has been a member of a street gang (0=no, 1=yes)	MCS	From a single item in age 14 questionnaire.
gang17	**Gang involvement age 17.** A binary variable that indicates if the individual currently at age 17 or in the past has been a member of a street gang (0=no, 1=yes)	MCS	From single item in age 17 questionnaire.
gang 1417	**Gang involvement age 14/17.** A binary variable that indicates if the individual currently at age 14 or 17 or in the past has been a member of a street gang (0=no, 1=yes)	MCS	Derived from variables: gang14 and gang17
ACEs	**Adverse Childhood Experiences (ACEs).** A continuous variable that summarises the number of ACEs during childhood (9 months to 11 years).	MCS	The exact items or variables making up this variable are still to be specified but initial ideas are outlined below in table 3.2.
PCEs	**Positive Childhood Experiences (PCEs).** A continuous variable that summarises the number of PCEs during childhood (9 months to 11 years).	MCS	The exact items or variables making up this variable are still to be specified but initial ideas are outlined below in table 3.2.
areacrim	**Neighbourhood violent crime.** A continuous	Police Recorded Crime (police.uk)	The category of crimes labelled

	measure of the per capita number of violent crimes reported to the police in each neighbourhood (LSOAs) of participants in the MCS.		violent crime will used.
IMDcrime	**Neighbourhood crime (England).** A measure of the per capita number of recorded crime in four major crime types (violence, burglary, theft, criminal damage). LSOA ranks range from 1 (most deprived) to 32482 (least deprived).	Index of Multiple Deprivation (England)	The crime domain will be used.
IMDsafety	**Neighbourhood crime (Wales).** The measure assesses the level of disadvantage that comes with living in an insecure neighbourhood, considering various factors like adult offenders, burglary, criminal damage, fire incidents, theft, violent crime, and the percentage of youth offenders. LSOA ranks range from 1 (most deprived) to 1896 (least deprived).	Index of Multiple Deprivation (Wales)	The Community Safety domain will be used.
LSOA	**Lower Layer Super Output Area (LSOA).** A categorical variable that identifies the LSOA or	MCS and Police Recorded Crime (Home Office)	LSOAs in the MCS is derived by the Centre for Longitudinal Studies from the

	neighbourhood where the MCS participants lives, and where violent crimes recorded to the police have taken place. This variable is used for linking the two datasets.		postcodes of participants home addresses. We will use LSOAs at the age 11 sweep, or the nearest previous sweep if non-response at age 11. LSOAs codes are included in both sources of the Police Recorded Crime data which will enable linkage.
eduh	**Household education.** A categorical variable that specifies the highest educational level in the household (No qualifications, NVQ1, NVQ2, NVQ3, NVQ4, NVQ5)	MCS	Derived from the initial sweep of the MCS using data from the main parent questionnaire in combination with the partner questionnaire (if relevant).
inch	**Household income.** A continuous variable that measures family financial resources during childhood (9 months to 11 years).	MCS	Derived from income information collected in first five sweeps of the MCS (age 9 months to 11 years).
IMDm	**Neighbourhood deprivation.** A modified measure of the Index of Multiple Deprivation (IMD) which excludes the crime domain. This is a continuous measure with higher values	MCS	Data on geographical deprivation has been linked to the MCS and is available for each survey sweep. We will use the age 11 sweep in our analyses, or the

	indicating a higher level of deprivation.		nearest previous sweep if non response at age 11.
matage	**Maternal age at birth**. This is age in years so a continuous variable.	MCS	The mother reported the month and year of their own birth and of the cohort child. Maternal age in years was then derived from these pieces of information rounding down to the closest year.
singpar	**Single parent status.** This is a continuous variable adding up the number of survey sweeps the main parent was not cohabiting with a partner from 9 months to age 11 (5 sweeps in total).	MCS	The main parent reported who lived in the household at each sweep.
eth	**Ethnicity of the cohort child.** Categorical variable (1.White, 2.Mixed, 3.Indian, 4.Pakistani & Bangladeshi, 5.Black or Black British, 6.Other)	MCS	Reported by the main parent in the initial sweep at age 9 months.
male	**Biological sex at birth**. Binary variable (1=male, 0=female)	MCS	Reported by the main parent in the initial sweep at age 9 months.

3.2. Measurement of key concepts

Table 3.2 Measurement of key concepts

Concept	How the concept will be measured and encoded

Adverse childhood experiences (ACEs)	These are experiences in childhood (age 9 months to age 11 years) which are thought to be risk factors for adverse outcomes. In terms of how we will measure ACEs, we will consider the framework from previous research using the MCS sample, which have included the following dimensions: 1) verbal maltreatment, 2) physical maltreatment, 3) parental divorce, 4) parental mental illness, 5) parental alcohol use, 6) parental drug use, and 7) domestic violence.[8 31]. However, we will explore the option of including further adversities as we plan to cover a longer period in childhood than in previous studies. The number of adverse experiences will be summed, so this is a continuous measure. In our analyses, a categorical measure will most likely be used that combines individuals with a high level of ACEs (e.g. 1.none, 2.one, 3.two, 4.three, 5. four or more).
Positive childhood experiences (PCEs)	These are experiences in childhood (age 9 months to age 11 years) which are thought to protect against adverse outcomes. In our measurement of PCEs we will draw on previous literature which has included aspects such as 1) good family communication, 2) positive relationships to trusted adults, 3) predictable home routines, 4) school connectedness, 5) positive peer influence support, 6) community involvement, 7) safe neighbourhood, 8) supportive neighbourhood.[9 14 15 32] To our knowledge no previous study has examined PCEs in the MCS. We will identify measures in the MCS that closely relate to these dimensions as well as consider additional positive aspects that may be relevant. The number of positive experiences will be summed, so this is a continuous measure. In our analyses, a categorical measure will most likely be used that combines individuals with a high level of PCEs (e.g. 1.none, 2.one, 3.two, 4.three, 5. four or more).
Neighbourhood violent crime	We will use police recorded crimes from police.uk and will use the violent crime category of this data. The total number of recorded crimes will be summed up for each Lower Layer Super Output Area (LSOA) which will define the neighbourhood. This will then by divided by the LSOA population estimate to obtain crime per capita. Checks on the appropriateness of using police.uk crime measures will be carried out by examining the consistency of the measure with IMD crime measures and Home Office reported crime. Our proposed measures of neighbourhood violent crime will be presented in the interim report.

Youth violence: *Assault*	Assault is defined as having pushed/shoved/hit/slapped/punched someone in the last 12 months either at age 14 or at age 17. • Age 14 item: In the last 12 months have you pushed or shoved**/**hit/slapped/punched someone? (yes/no) • Age 17 item: In the last 12 months have you pushed or shoved**/**hit/slapped/punched someone? (yes/no) The measure used for analysis is binary (1=yes/0=no). A positive response to either of these items =1, if both are negative =0.
Youth violence: *Weapons involvement*	Weapons involvement is defined as: a) having carried a weapon ever at age 14 or in the past year at age 17. OR b) having used a weapon in the last 12 months at age 14 or 17. • Age 14 item: In the last 12 months have you used or hit someone with a weapon? (yes/no) • Age 14 item: Have you ever carried a knife or other weapon for your own protection because someone else asked you to or in case you get into a fight? (yes/no) • Age 17 item: In the last 12 months have you hit someone with or used a weapon? • Age 17 item: In the last 12 months have you carried a knife or other weapon? For your own protection, because someone else asked you to or in case you get into a fight. The measure used for analysis is binary (1=yes/0=no). A positive response to any of these items =1, if all are negative =0.
Youth violence: *Gang involvement*	Gang involvement is defined as being a member of a street gang currently or in the past, either at age 14 or 17 • Age 14 and 17 item: Are you a member of a street gang? (By a street gang, we mean groups of young people who hang around together and: have a specific area or territory; have a name, a colour or something else to identify the group; possibly have rules or a leader; who may commit crimes together. (no, yes currently, yes in the past) The measure used for analysis is binary (1=yes/0=no). A positive response to either of these items =1, if both are negative =0.

3.3. Missing data and attrition

Handling missing data in the MCS

Examining extent of missing data

As in all longitudinal studies, there is attrition over time, which disproportionately tends to affect the more disadvantaged participants, and, if not addressed correctly, can bias results. Of the 19,519 children who were initially recruited for the study, 10,757 cohort members provided data at age 17, with even lower response rates on some specific survey questions. Those involved in violence are highly likely to attrit at a higher rate than others, which we will directly investigate using data from earlier sweeps and examining their correlation with age 14 and age 17 participation. Specifically, antisocial behaviour at age 11, conduct problems across childhood (age 3-11), and also family socioeconomic status will be examined in relation to missing data.

Multiple imputations and weighting

To deal with missing data in the MCS, we will use multiple imputations. This is essentially a way of 'restoring' missing data, under certain plausible assumptions, and it is an efficient method for replicating population estimates in longitudinal data when data are missing.[33]

Multiple datasets will be imputed using chained equations, whereby missing data are 'filled' in each dataset using information from all other predictor and outcome variables in the analytical models. The multiple imputations will further include several auxiliary variables, which are not included in the analytical models, but are correlated with attrition. The substantial analyses will then be carried out, producing a single combined estimate across the multiple datasets.

In addition to multiple imputations, survey weights will be used in all analyses to account for the complex sampling strategy used in the initial MCS survey. Using these weights and the imputation strategy will allow us to provide measures and estimates that are as close as possible to being nationally representative of this generation of young people.

Missing variables

Although MCS data is extremely rich and detailed, there are nevertheless some aspects included in the original ACEs framework that are not measured in this cohort study. These are 1) sexual abuse and 2) household criminality (prison sentence). As in previous studies that have examined ACEs using the MCS [8 31 34], we have no option but to omit these dimensions in our ACE measure and acknowledge it as a study limitation.

3.4. Other sources of bias

Omitted variable and selection bias

Because the study design is inherently observational and not a controlled experiment, we can not estimate causal relationships. However, we can greatly reduce the extent of omitted

variables bias by controlling for an extensive set of variables in the regression analysis, owing to the richness of MCS data collected over time, and the additional linkage with area-level crime data. By reducing to the extent possible omitted variables bias, our estimates are as close to causal as we can get in this case and using observational data.

Representation of ethnic minority and disadvantaged families

From the outset the MCS was carefully designed to ensure adequate representation of disadvantaged families and ethnic minority children. Areas with a high concentration of ethnic minorities, and also areas high on the Child Poverty Index were oversamples, details of which can be found elsewhere. [35] Measures taken to reach ethnic minority groups included letters and information leaflets translated into the seven most common non-English languages. However, the extent to which the respondents coming forward for participation fully represent disadvantaged families and ethnic minorities is difficult to fully assess. It is likely that ethnic minority families who felt the most marginalised were less likely to agree to participation. It is also likely to ethnic minority participant are more likely to drop out of the study over time and have a higher proportion of missing data on the youth violence outcomes. We are able to directly examine such patterns of attrition/missingness in our data. As detailed above we will use multiple imputation to mitigate any bias that attrition/missingness can cause, including that by ethnic minority participants. In our fully imputed or 'restored' sample, we will examine whether the frequency of ethnic minority group match that of 17 year olds in the general population. In our proposed analyses we will explicitly examine whether results (how predictors are related to youth violence) vary by ethnic group. However, it is possible that we may not be able to fully examine each ethnic group, due to difficulties running analyses on the multiple imputed datasets in cases where outcomes, predictors, or control variables have no observations in some datasets. Modifications are likely to include collapsing categories in some variables, such as using a binary measure for ethnic (i.e. ethnic minority vs white).

4. About the analysis

4.1. Overview of analytical approach

Prior to any analyses that directly address the substantial research questions of the study, exploratory work will be carried out to develop appropriate exposure variables, i.e. ACEs, PCEs and neighbourhood violent crime. As outlined further above, the plan is to use police.uk provided data in our measure of neighbourhood violent crime. However, we will need to carefully examine the appropriateness of using this data, which is the objective in research questions 4. The focus of the interim report will be on the development of variables, and checking of data, as well as initial results for the first four research questions.

Following this interim work, analyses will focus directly on addressing the full research questions of the study. First the sample and data will be described, which will include descriptive statistics and characteristics of the final (imputed) study sample, the main outcomes, exposure variables, and control variables, as well as bivariate analyses between these study variables.

This will be followed by the more substantial analyses addressing research questions 1-3 and 5-10. For these examinations we will use multivariate logistic regressions to examine to what extent youth violence at age 14/17 (assault, weapons involvement, gang involvement) is predicted by ACEs and PCEs, neighbourhood violent crime, and the intersection between these. For the examinations of the intersection between the exposure variables (research questions 3, 6, 7, 8, 9 and 10), a split sample approach will be taken and if data allows it interactions between variables will be included in regressions.

Multivariate regression analyses will include variables that may confound the association between exposures and outcomes. For the analysis examining ACEs and PCEs as exposures, covariates will include parental education and household income. For the analyses examining neighbourhood crime as the exposure, covariates will include a modified measure of area deprivation that excludes the crime domain, and also family level variables (ACEs, PCEs, maternal age, single parent status, household income, parental education) will be controlled for as these may drive selection into neighbourhood. The inclusion of covariates will strengthen the robustness of findings by providing adjusted associations between our exposures (ACEs, PCEs, neighbourhood crime) and our youth violence outcomes (assault, weapons involvement, gang involvement), rather than these associations arising because of other factors. To examine the fit of the models, the pseudo R-square statistic will be reported which in logistic regression indicates the extent to which the independent variable (youth violence) is explained by variables in the model.

A clear advantage of the Multivariate OLS regression to address our research question over other techniques such as Multilevel Modelling (MLM) methods (also referred to as Random Effects models) is that it allows us to lessen potential omitted variable biases by controlling for area-level fixed effects, e.g., Police Force fixed effects, while also allowing the fixed effects to be correlated with the error term. Additionally, the scattered geographical distribution of MCS participants is not conducive to modelling hierarchical data structures using MLM.

In our analyses we will additionally examine if results vary by gender or ethnicity, which will be done by running regression models separately for males and females, and for white and ethnic minority participants, and interactions between these moderators and exposures will also be examined. Only data from England and Wales will be used in analyses as specified in the YEF research scope.

4.2. Approach to addressing each research question

Research question 1

Research question	**Which single and cumulative ACEs are associated with youth violence, and what are the magnitudes of the associations?**
Hypothesis, if relevant	We expect single and cumulative ACEs to be associated with a higher likelihood of youth violence.
What will you be able to say by the interim report	Preliminary results will be shown in tables or graphs and briefly summarised.
Descriptive analysis, if relevant	Each single and cumulative ACEs measure will be described in terms of frequency. These descriptives will be provided for the sample overall and by each youth violence outcome, as well as by gender and ethnicity, parental education, household income, single parent status, and maternal age.
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons involvement, gang involvement), all of which are binary measures. Each single ACEs measure and the cumulative ACEs measure, respectively, will be treated as the predictor of interest in these models. Covariates will include parental education, household income, single parent status, and maternal age.
Estimating equation, if relevant	$\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Single or cumulative ACEs X_2 = Parental education X_3 = Household income X_4 = Single parent status X_5 = Maternal age at birth

	The coefficient of interest is $b_{1,}$ which is the estimate for the association between the exposure (single and cumulative ACESs) and youth violence.
What does the approach need to succeed (constraints/assumptions)?	We use multiple imputed data and a main assumption is that for each dataset there are observations for each violence outcome across the predictor variable, control variables and also the moderator variables (gender and ethnicity).
Uncertainty and inference	95% confidence intervals will be reported, which will be our threshold for statistical significance.
Robustness checks	1. We will include area violent offending and area deprivation as additional covariates. The reason for not including these in the main analyses is that they are likely to be mediators rather than confounders. For example, a high level of ACEs is unlikely to be caused by these neighbourhood level characteristics. Rather, families characterised by high ACEs and dysfunction is likely to influence what neighbourhood they live in. Including these additional covariates will likely reduce the estimated associations between ACEs and youth violence, but if the association still holds statistically it provides additional strength to our findings. 2. We will run analyses on non-multiple imputed data to examine any differences
Subgroup you intend to study	We intend to analyse the association between the cumulative ACEs measure and youth violence outcomes by gender (males and females) and by ethnic groups (5 categories). We plan to do this using a split-sample approach by running separate models for these subgroups.
Changes to the analysis	Modifications to the analyses will be made if there are difficulties running analyses on the multiple imputed datasets. This can arise because of outcomes, predictor, or

	control variables having small cell numbers, and in some datasets there may be no observations, in which case analyses will not run. Modifications are likely to include collapsing categories in some variables, such as using a binary measure. A binary approach may be taken to interaction analyses generally across the study to avoid small categories (i.e. ACEs, PCEs and neighbourhood violent crime dichotomised as high vs low). This approach may also aide interpretability of results.

Research question 2

Research question	**Which single and cumulative PCEs are associated with youth violence, and what are the magnitudes of the associations?**
Hypothesis, if relevant	We expect single and cumulative PCEs to be associated with a lower likelihood of youth violence.
What will you be able to say by the interim report	Preliminary results will be shown in tables or graphs and briefly summarised.
Descriptive analysis, if relevant	Each single and cumulative PCEs measure will be described in terms of frequency. These descriptives will be provided for the sample overall and by each youth violence outcome, as well as by gender and ethnicity, parental education, household income, single parent status, and maternal age.
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons involvement, gang involvement), all of which are binary measures. Each single PCEs measure and the cumulative PCEs measure, respectively, will be treated as the predictor of interest in these models. Covariates will include parental education, household income, single parent status, and maternal age.

Estimating equation, if relevant	$\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Single or cumulative PCEs X_2 = Parental education X_3 = Household income X_4 = Single parent status X_5 = Maternal age at birth The coefficient of interest is $b_{1,}$ which is the estimate for the association between the exposure (single and cumulative PCESs) and youth violence.
What does the approach need to succeed (constraints/assumptions)?	Same as for question 1
Uncertainty and inference	Same as for question 1
Robustness checks	Similar approach as for question 1
Subgroup you intend to study	Similar approach as for question 1
Changes to the analysis	Similar approach as for question 1

Research question 3

Research question	**Do PCEs attenuate the association between ACEs and youth violence?**
Hypothesis, if relevant	We hypothesise that a higher level of cumulative PCEs will attenuate the association between cumulative ACEs and youth violence.

What will you be able to say by the interim report	Preliminary results will be shown in tables or graphs and briefly summarised.
Descriptive analysis, if relevant	Not applicable
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be use as described for question 1, with a split sample model for high versus low levels of PCEs. In addition, a model will be run with an interaction term between ACEs and PCEs.
Estimating equation, if relevant	Equation for each level of PCEs (split sample) is the same as for question 1 above. Below is the model with the interaction term. $\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_1X_2+b_4X_4+b_5X_5+b_6X_6+b_7X_7$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Cumulative ACEs X_2= Cumulative PCEs X_3= Interaction between ACEs and PCEs X_4 = Parental education X_5 = Household income X_6 = Single parent status X_7 = Maternal age at birth The coefficient of interest is $b3$, which is the interaction between ACEs and PCESs and its association with youth violence.
What does the approach need to succeed (constraints/assumptions)?	We use multiple imputed data and a main assumption is that for each dataset there are observations for each violence outcome across the predictor variable, control

	variables and also the moderator variables (gender and ethnicity). Another assumption is that the sample has enough statistical power to detect interaction effects, which requires more power than main effects.
Uncertainty and inference	Same as for question 1
Robustness checks	Similar approach as for question 1
Subgroup you intend to study	Not applicable
Changes to the analysis	Similar approach as for question 1

Research question 4

Research question	**To what extent is neighbourhood violent crime using police.uk data a valid and reliable measure?**
Hypothesis, if relevant	We expect LSOAs to have a similar ranking in terms of level of neighbourhood crime measured using police.uk data as in the IMD data. We expect the police.uk measure of neighbourhood crime to show the same trend over time as the published Home Office national figures for crime.
What will you be able to say by the interim report	Recommendations will be made with regards to the reliability of police.uk data.
Descriptive analysis, if relevant	Descriptive statistics of the Index of Multiple Deprivation in England (2010) and in Wales (2011) will be compared with summary statistics of aggregated data form the police.uk data (2011-2013). Similarly, we will compare trends of levels of crime using Home Office statistics on police recorded violent crime and compared them with aggregated police.uk crime data.

Models, specifications and statistical techniques used, if relevant	Not applicable
Estimating equation, if relevant	Not applicable.
What does the approach need to succeed (constraints/assumptions)?	We expect that some restrictions on geographical and time coverage of IMD, police.uk and Home Office data will be needed to facilitate comparations.
Uncertainty and inference	Not applicable.
Robustness checks	Not applicable.
Subgroup you intend to study	Not applicable.
Changes to the analysis	Not applicable.

Research question 5

Research question	**Are rates of violent crime in one's neighbourhood associated with youth violence?**
Hypothesis, if relevant	We hypothesise that a higher level of violent crime in the neighbourhood is associated with a higher likelihood of youth violence.
What will you be able to say by the interim report	Recommendations will be made with regards to which area crime variables and data source should be used.
Descriptive analysis, if relevant	The overall measure of neighbourhood violent crime will be summarised (mean, sd, max, min), and frequency of its constituent parts (crime categories) will also be presented. In addition, the mean of neighbourhood violent crime will be provided by each youth violence outcome, as well as by gender and ethnicity, parental education, household income, single parent status, maternal age, neighbourhood deprivation, and ACEs and PCEs.
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons involvement, gang involvement), which are all binary measures. Neighbourhood violent crime will be treated as the main predictor variable in these models. Covariates will include parental education, household income, single parent status, maternal age, and ACEs and PCEs.
Estimating equation, if relevant	$\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+b_7X_7$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Neighbourhood violent crime X_2 = Parental education X_3 = Household income X_4 = Single parent status

	X_5 = Maternal age at birth X_6 = ACEs X_7 = PCEs The coefficient of interest is $b_{1,}$ which is the estimate for the association between the neighbourhood violent crime and youth violence.
What does the approach need to succeed (constraints/assumptions)?	Same as for question 1
Uncertainty and inference	Same as for question 1
Robustness checks	1. We will include a measure of area deprivation as an additional covariate. This will establish the prediction estimate for neighbourhood level violence, over and above area-based deprivation, which is likely to be highly correlated. 2. We will run analyses on non-multiple imputed data to examine any differences
Subgroup you intend to study	Similar approach as for question 1
Changes to the analysis	Similar approach as for question 1

Research question 6

Research question	**Does the association between ACEs and youth violence differ for those in neighbourhoods with high versus low levels of violent crime?**
Hypothesis, if relevant	As a high level of violence in the neighbourhood is a potential additional risk factor, we hypothesise that ACEs have a stronger association with youth violence in

	neighbourhoods with high levels of violent crime compared to areas with lower levels.
What will you be able to say by the interim report	No results will be included in the interim report.
Descriptive analysis, if relevant	Not applicable
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons involvement, gang involvement), all of which are binary measures. The association between ACEs and youth violence will be examined for those living in high versus low crime neighbourhoods (split sample). In addition, an interaction between ACEs and neighbourhood violent crime will be added to the model. Covariates will include parental education, household income, single parent status, and maternal age.
Estimating equation, if relevant	The equation for the split sample analyses is the same as for question 1 above. Below is the model with the interaction term. $\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_1X_2+b_4X_4+b_5X_5+b_6X_6+b_7X_7$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Cumulative ACEs X_2= Neighbourhood violent crime X_3= Interaction between ACEs and neighbourhood violent crime X_4 = Parental education X_5 = Household income X_6 = Single parent status X_7 = Maternal age at birth

	The coefficient of interest is b_3, which is the interaction between ACEs and neighbourhood violent crime and its association youth violence.
What does the approach need to succeed (constraints/assumptions)?	Same as for question 3
Uncertainty and inference	Same as for question 1
Robustness checks	1. A model will be run that replaces neighbourhood violent crime with general neighbourhood deprivation. This will establish whether any interaction with area crime is different to the interaction with area deprivation. 2. We will run analyses on non-multiple imputed data to examine any differences
Subgroup you intend to study	Not applicable
Changes to the analysis	Similar approach as for question 1

Research question 7

Research question	**Does the association between PCEs and youth violence differ for those in neighbourhoods with high versus low levels of violent crime?**
Hypothesis, if relevant	We hypothesise that the association between PCEs and youth violence will be weaker in neighbourhoods with a high level of violent crime. This is because neighbourhood violent crime is a potential risk factor and may undermine the protectiveness of PCEs for youth violence.

What will you be able to say by the interim report	No results will be included in the interim report.
Descriptive analysis, if relevant	Not applicable
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons involvement, gang involvement), all of which are binary measures. The association between PCEs and youth violence will be examined for those living in high versus low crime neighbourhoods (split sample). In addition, an interaction between PCEs and neighbourhood violent crime will be added to the model. Covariates will include parental education, household income, single parent status, and maternal age.
Estimating equation, if relevant	The equation for the split sample analyses is the same as for question 2 above. Below is the model with the interaction term. $\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_1X_2+b_4X_4+b_5X_5+b_6X_6+b_7X_7$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Cumulative PCEs X_2= Neighbourhood violent crime X_3= Interaction between PCEs and neighbourhood violent crime X_4 = Parental education X_5 = Household income X_6 = Single parent status X_7 = Maternal age at birth The coefficient of interest is b_3, which is the interaction between PCEs and neighbourhood violent crime and its association youth violence.

What does the approach need to succeed (constraints/assumptions)?	Same as for question 3
Uncertainty and inference	Same as for question 1
Robustness checks	1. A model will be run that replaces neighbourhood violent crime with general neighbourhood deprivation. This will establish whether any interaction with area crime is different to the interaction with area deprivation. 2. We will run analyses on non-multiple imputed data to examine any differences
Subgroup you intend to study	Not applicable
Changes to the analysis	Similar approach as for question 1

Research question 8

Research question	**Do PCEs attenuate the association between ACEs and violent crime more in low-crime areas or in high-crime areas?**
Hypothesis, if relevant	Because neighbourhood violent crime is a potential risk factor, the extent to which PCEs reduce the association between ACEs and youth violence is hypothesised to be lower in areas with high levels of violent crime.
What will you be able to say by the interim report	No results will be included in the interim report.
Descriptive analysis, if relevant	Not applicable

Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons involvement, gang involvement), all of which are binary measures. The association between ACEs and violent crime will be examined for each combination of high and low PCEs and neighbourhood violent crime (split sample). In addition, a 3-way interaction between ACEs, PCEs and neighbourhood violent crime will be added to the model. Covariates will include parental education, household income, single parent status, and maternal age.
Estimating equation, if relevant	The equation for the split sample analyses is the same as for question 1 above. Below is the model with the interaction term. $\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_3+b_4X_1X_2X_3+b_5X_5+b_6X_6+b_7X_7+b_8X_8$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Cumulative ACEs X_2 = Cumulative PCEs X_3= Neighbourhood violent crime X_4= Interaction between ACEs, PCEs and neighbourhood violent crime X_5 = Parental education X_6 = Household income X_7 = Single parent status X_8 = Maternal age at birth The coefficient of interest is b_4, which is the interaction between PCEs, ACEs and neighbourhood violent crime, predicting youth violence.
What does the approach need to succeed	Same as for question 3

(constraints/assumptions) ?	
Uncertainty and inference	Same as for question 1
Robustness checks	1. A model will be run that replaces neighbourhood violent crime with general neighbourhood deprivation. This will establish whether any area crime moderates the association differently than area deprivation. 2. We will run analyses on non-multiple imputed data to examine any differences
Subgroup you intend to study	Not applicable
Changes to the analysis	Similar approach as for question 1

Research question 9

Research question	**Do ACEs amplify the association between neighbourhood violent crime and youth violence?**
Hypothesis, if relevant	Because ACEs are likely to be an additional risk factor, we hypothesise that neighbourhood violent crime will have a stronger association with youth violence for those with a high level of ACEs.
What will you be able to say by the interim report	No results will be included in the interim report.
Descriptive analysis, if relevant	Not applicable
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons

	involvement, gang involvement), all of which are binary measures. The association between neighbourhood violent crime and youth violence will be examined for those with high versus low level of ACEs (split sample). In addition, an interaction between ACEs and neighbourhood violent crime will be added to the model. Covariates will include parental education, household income, single parent status, and maternal age.
Estimating equation, if relevant	The equation for the split sample analyses is the same as for question 4 above. Below is the model with the interaction term. $\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_1X_2+b_4X_4+b_5X_5+b_6X_6+b_7X_7$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Cumulative ACEs X_2= Neighbourhood violent crime X_3= Interaction between ACEs and neighbourhood violent crime X_4 = Parental education X_5 = Household income X_6 = Single parent status X_7 = Maternal age at birth The coefficient of interest is b_3, which is the interaction between ACEs and neighbourhood violent crime and its association youth violence.
What does the approach need to succeed (constraints/assumptions)?	Same as for question 1
Uncertainty and inference	Same as for question 1

Robustness checks	1. We will include a measure of area deprivation as an additional covariate. This will establish the prediction estimate for neighbourhood level violence, over and above area-based deprivation, which is likely to be highly correlated. 2. We will run analyses on non-multiple imputed data to examine any differences
Subgroup you intend to study	Not applicable
Changes to the analysis	Similar approach as for question 1

Research question 10

Research question	**Do PCEs attenuate the association between neighbourhood violent crime and youth violence?**
Hypothesis, if relevant	Because PCEs are likely a protective factor, we hypothesise that neighbourhood violent crime will have a weaker association with youth violence for those with a high level of PCEs.
What will you be able to say by the interim report	No results will be included in the interim report.
Descriptive analysis, if relevant	Not applicable
Models, specifications and statistical techniques used, if relevant	Logistic regression methods will be used to examine each of the youth violence outcomes (assault, weapons involvement, gang involvement), all of which are binary measures. The association between neighbourhood violent crime and youth violence will be examined for those with high versus low level of PCEs (split sample). In addition, an interaction

	between PCEs and neighbourhood violent crime will be added to the model. Covariates will include parental education, household income, single parent status, and maternal age.
Estimating equation, if relevant	The equation for the split sample analyses is the same as for question 4 above. Below is the model with the interaction term. $\ln[Y/(1-Y)]=a+b_1X_1+b_2X_2+b_3X_1X_2+b_4X_4+b_5X_5+b_6X_6+b_7X_7$ Y = Youth violence (assault, weapons involvement, gang involvement) a = Constant X_1 = Cumulative PCEs X_2= Neighbourhood violent crime X_3= Interaction between PCEs and neighbourhood violent crime X_4 = Parental education X_5 = Household income X_6 = Single parent status X_7 = Maternal age at birth The coefficient of interest is b_3, which is the interaction between PCEs and neighbourhood violent crime and its association youth violence.
What does the approach need to succeed (constraints/assumptions)?	Same as for question 3
Uncertainty and inference	Same as for question 1
Robustness checks	1. We will include a measure of area deprivation as an additional covariate. This will establish the prediction estimate for neighbourhood level violence, over and above area-based deprivation, which is likely to be highly correlated.

	2. We will run analyses on non-multiple imputed data to examine any differences
Subgroup you intend to study	Not applicable
Changes to the analysis	Similar approach as for question 1

5. Project management

5.1. Risks and mitigations

Table 5.1 Risks and mitigations

Number	Risk	Likelihood (*Low/Medium/ High)*	Mitigation
1	Fail to obtain permission to link MCS data with crime data.	Low	We have already obtained permission for another similar project that links MCS to police.uk data. Should we however fail to obtain this permission we will drop the neighbourhood violent crime aspect of the project. We would instead offer to do additional analyses using just the MCS on other dimensions of interest to the YEF.
2	Project research staff may leave, and recruitment may cause delays	Low	We would look to recruit ASAP. Ideally from our own centre which has a number of researchers who are highly experienced in working on large-scale longitudinal cohort data.

5.2. Timeline

Table 5.2 Timeline

Date	Activity	Staff responsible/leading
Feb 2023	**Application for research ethical approval** • Ethical approval from the UCL IoE Research Ethics Committee will be sought.	AV
Mar 2023	**Development of research study plan** • The initial research study plan as required by the YEF will be developed, which will specify the agreed research questions and provisional methodological approach.	AV
Mar-April 2023	**Permissions for data access and linkage** • Application to the Home Office for data sharing of police recorded crime. • Application for data linkage through the Centre for Longitudinal Studies Data Access Committee (CLSDAC). • Registration with the Safe Data Haven and staff to undergo training in using this platform.	NL
April-July 2023	**Initial examination of police.uk crime data** • Examination of consistency of police.uk reported violent crime with IMD crime domain/community safety. • Examination of consistency across time police.uk reported violent crime at national level with Home Office • Linking of police.uk crime data to the MCS (unimputed data)	NL
April–July 2023	**Preparing the MCS dataset** • Preparation of the MCS dataset specific to the project, which will involve merging data across surveys, and examining and creating study variables (outcomes, exposures,	AV

	control variables, auxillary imputation variables).	
Sep 2023	**Multiple imputations of MCS data** • Multiple imputation of MCS datasets to deal with missing data due to attrition. • Variables constructed following multiple imputation	AV
Oct 2023	**Analyses for descriptives and research questions 1-3** • Descriptive analyses of sample and study variables • Analyses addressing research questions 1, 2, and 3.	AV
Nov 2023 – Dec 2024	**Writing up of interim report results** • Initial examinations and checks on variables. • Recommendations made for variables and data sources • Descriptive analyses of sample and study variables • Initial findings (tables or graphs and brief summary) related to research questions 1, 2, and 3. • Comments from YEF responded and decisions made for research approach	AV, NL
Jan – Mar 2024	**Analyses for research questions 5-10** • Merging multiple imputed MCS data with area crime data • Analyses addressing research questions 5 to 10.	AV, NL, EF
April - June 2024	**Write up of final report** • Presentation of full and final results in graphs and tables • Write up of full and final results • Write up all other sections of report (methods, discussion, and interpretation of results)	AV, NL, EF

July 2024	**Submission of final report to YEF**	AV, NL, EF

6. References

1. Youth Justice Board/Ministry of Justice. Youth Justice Statistics 2019/2020: England and Wales, 2021.

2. Derzon JH. The correspondence of family features with problem, aggressive, criminal, and violent behavior: a meta-analysis. *Journal of Experimental Criminology* 2010;6(3):263-92. doi: 10.1007/s11292-010-9098-0

3. Farrington DP. Families and crime. In: Wilson JQ, Petersilia J, eds. Crime and public policy. New York: Oxford University Press 2011:130-57.

4. Appleton AA, Holdsworth E, Ryan M, et al. Measuring Childhood Adversity in Life Course Cardiovascular Research: A Systematic Review. *Psychosom Med* 2017;79(4):434-40. doi: 10.1097/PSY.0000000000000430 [published Online First: 2016/11/29]

5. Asmussen K, F F, Drayton E, et al. Adverse childhood experiences: What we know, what we don't know, and what should happen next: Early Intervention Foundation, 2020.

6. Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *Am J Prev Med* 1998;14(4):245-58. doi: 10.1016/s0749-3797(98)00017-8 [published Online First: 1998/06/23]

7. Houtepen LC, Heron J, Suderman MJ, et al. Adverse childhood experiences: associations with educational attainment and adolescent health, and the role of family and socioeconomic factors. Analysis of a prospective cohort study. *bioRxiv* 2019:612390.

8. Straatmann VS, Lai ET, Taylor-Robinson DC. How do adverse childhood experiences explain social inequalities in adolescent health outcomes? findings from the UK millennium cohort study. *The Lancet* 2018;392:S83.

9. Bethell C, Jones J, Gombojav N, et al. Positive Childhood Experiences and Adult Mental and Relational Health in a Statewide Sample: Associations Across Adverse Childhood Experiences Levels. *JAMA Pediatr* 2019;173(11):e193007. doi: 10.1001/jamapediatrics.2019.3007 [published Online First: 2019/09/10]

10. Crandall A, Miller JR, Cheung A, et al. ACEs and counter-ACEs: How positive and negative childhood experiences influence adult health. *Child Abuse & Neglect* 2019;96:104089. doi: 10.1016/j.chiabu.2019.104089

11. Crouch E, Radcliff E, Merrell MA, et al. Positive Childhood Experiences Promote School Success. *Matern Child Health J* 2021;25(10):1646-54. doi: 10.1007/s10995-021-03206-3 [published Online First: 2021/08/15]

12. Sege RD, Harper Browne C. Responding to ACEs With HOPE: Health Outcomes From Positive Experiences. *Acad Pediatr* 2017;17(7S):S79-S85. doi: 10.1016/j.acap.2017.03.007 [published Online First: 2017/09/04]

13. Hillis SD, Anda RF, Dube SR, et al. The Protective Effect of Family Strengths in Childhood against Adolescent Pregnancy and Its Long-Term Psychosocial Consequences. *Perm J* 2010;14(3):18-27. doi: 10.7812/tpp/10-028 [published Online First: 2010/09/17]

14. Baglivio MT, Wolff KT. Positive childhood experiences (PCE): Cumulative resiliency in the face of adverse childhood experiences. *Youth violence and juvenile justice* 2021;19(2):139-62.

15. Craig JM, Baglivio MT, Wolff KT, et al. Do social bonds buffer the impact of adverse childhood experiences on reoffending? *Youth violence and juvenile justice* 2017;15(1):3-20.

16. Qu GB, Ma SD, Liu HX, et al. Positive childhood experiences can moderate the impact of adverse childhood experiences on adolescent depression and anxiety: Results from a cross-sectional survey. *Child Abuse & Neglect* 2022;125:105511. doi: 10.1016/j.chiabu.2022.105511

17. Sampson RJ. Neighborhood effects, causal mechanisms and the social structure of the city. In: Demeulenaere P, ed. Analytical sociology social mechanisms. New York: Cambridge University Press 2011:227-49.

18. Centre for Crime and Justice Studies. Young people, and gun and knife crime: a review of the evidence. London: King's College London, 2009.

19. Ham Mv, Manley D, Bailey N, et al. Neighbourhood effects research: New perspectives. Neighbourhood effects research: New perspectives: Springer 2012:1-21.

20. Wikström POH, Loeber R. Do disadvantaged neighborhoods cause well-adjusted children to become adolescent delinquents? A study of male juvenile serious offending, individual risk and protective factors, and neighborhood context. *Criminology* 2000;38(4):1109-42.

21. Bellis MA, Lowey H, Leckenby N, et al. Adverse childhood experiences: retrospective study to determine their impact on adult health behaviours and health outcomes in a UK population. *Journal of Public Health* 2014;36(1):81-91. doi: 10.1093/pubmed/fdt038

22. Duke NN, Pettingell SL, McMorris BJ, et al. Adolescent violence perpetration: associations with multiple types of adverse childhood experiences. *Pediatrics*

2010;125(4):e778-86. doi: 10.1542/peds.2009-0597 [published Online First: 2010/03/17]

23. Fox BH, Perez N, Cass E, et al. Trauma changes everything: Examining the relationship between adverse childhood experiences and serious, violent and chronic juvenile offenders. *Child Abuse & Neglect* 2015;46:163-73. doi: 10.1016/j.chiabu.2015.01.011

24. Leban L. The effects of adverse childhood experiences and gender on developmental trajectories of internalizing and externalizing outcomes. *Crime and Delinquency* 2021;67(5):631-61.

25. Leban L, Gibson CL. The role of gender in the relationship between adverse childhood experiences and delinquency and substance use in adolescence. *Journal of criminal justice* 2020;66:101637.

26. Kroneman L, Loeber R, Hipwell AE. Is neighborhood context differently related to externalizing problems and delinquency for girls compared with boys? *Clinical child and family psychology review* 2004;7:109-22.

27. Jones MS, Shurtliff TM, Gerlinger J. Racial differences in the effects of early adverse childhood experiences on adolescent delinquency: The role of middle childhood mechanisms. *Journal of Criminal Justice* 2022;82:101977.

28. Villadsen A, Fitzsimons, E. Prevalence and predictors of weapon carrying and use and other offences at age 17: Evidence from the UK Millennium Cohort Study. London: UCL Centre for Longitudinal Studies, 2021.

29. Office for National Statistics. Lower layer Super Output Area population estimates (supporting information) 2021 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates.

30. UK Data Service. Millennium Cohort Study, 2018.

31. Jackson DB, Jindal M, Testa A, et al. Adverse childhood experiences and adolescent police contact in the United Kingdom. *Pediatrics* 2022;150(4)

32. Narayan AJ, Rivera LM, Bernstein RE, et al. Positive childhood experiences predict less psychopathology and stress in pregnant women with childhood adversity: A pilot study of the benevolent childhood experiences (BCEs) scale. *Child abuse and neglect* 2018;78:19-30.

33. Mostafa T, Narayanan M, Pongiglione B, et al. Missing at random assumption made more plausible: evidence from the 1958 British birth cohort. *J Clin Epidemiol* 2021;136:44-54. doi: 10.1016/j.jclinepi.2021.02.019

34. Nweze T, Ezenwa M, Ajaelu C, et al. Childhood mental health difficulties mediate the long-term association between early-life adversity at age 3 and poorer cognitive

functioning at ages 11 and 14. *Journal of Child Psychology and Psychiatry* 2023 doi: doi:10.1111/jcpp.13757

35. Plewis I, Calderwood L, Hawkes D, et al. Millennium Cohort Study: technical report on sampling. London: Centre for Longitudinal Studies, 2007.