# The ‘where?’ and ‘when?’ of children and young people’s involvement in violence

## Secondary Data Project Analysis Plan

Evaluating institution: Nottingham Trent University
Principal investigator: Professor Andy Newton

---

## Analysis Plan for YEF Secondary Data Analysis Projects

### Project summary

| | |
|---|---|
| **Project title** | The 'where?' and 'when?' of children and young people's involvement in violence |
| **Research Team** | Nottingham Trent University |
| **Principal investigator** | Professor Andy Newton |
| **Analysis plan author(s)** | Andy Newton, James Hunter, Rich Pickford, Becky Thompson, Erin Gibson |
| **Overarching research question** | Where and when are children and young people most likely to be involved in violence, and which local contextual factors best predict these concentrations of violence? |
| **Supporting research question(s)** | (i) Where are the highest concentrations of violence involving children and young people?<br>(ii) When do these concentrations of violence involving children and young people occur?<br>(iii) What local contextual factors best explain the location and timing concentrations of violence involving children and young people? |
| **Dataset(s) to be used** | Police-recorded crime data; Ordnance Survey (OS) datasets (Open Roads, Points of Interest, Open Greenspace); Office for National Statistics (ONS) Census 2021; MHCLG Indices of Deprivation (2025); University College London (UCL) Space Syntax; Nottingham Trent University (NTU) Community Engagement Area Classification; Police Powers and Procedures: Stop and Search |
| **Population characteristics** | Young people (11-25) who are offenders or victims of violence |
| **Years data spans** | April 2022 to March 2025 (3 years) |

| | |
|---|---|
| **Geographic coverage** | England and Wales: This research will utilise case studies of three to five police force areas – to be specified once data sharing contracts are agreed with force areas. |
| **Primary outcome(s) investigated (if relevant)** | Youth violence (police crime flagged by offender and or victim aged under 25) |
| **Main method(s) to be used or tested** | Descriptive study using regression modelling |

## Analysis plan history

| Version | Date | Reason for revision |
|---|---|---|
| 1.X [*latest*] | | |
| 1.1 | 19/11/2025 | Revisions in response to reviewer feedback |
| 1.0 [*original*] | 13/10/2025 | |

*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. About the project ..... 1**

*1.1. Background to the project ..... 1*

*1.3. Research question(s) ..... 1*

*1.4. Hypotheses ..... 2*

*1.5. Key concepts ..... 7*

**2. About the datasets ..... 10**

*2.1. Overview of datasets used ..... 10*

*2.2. Primary Data Sources ..... 10*

*2.3. Secondary data source(s) ..... 11*

*2.4. Access and data protection ..... 16*

*2.5. Ethics ..... 19*

**3. Methods ..... 21**

*3.1. Research design ..... 21*

*3.2. Participants ..... 21*

*3.3. Variables and measurement ..... 22*

***Other measures*** *..... 23*

***3.4. Sample size*** *..... 24*

***3.5. Research methods and analysis*** *..... 24*

*3.6. Missing data ..... 27*

*3.8. Other potential limitations or sources of bias ..... 28*

**4. Project management ..... 30**

*4.1. Risks and mitigations ..... 30*

*4.2. Timeline ..... 32*

**5. References ..... 34**

# 1. About the project

## 1.1. Background to the project

It is well established that crime including violence is highly concentrated, both geographically and by time of day. Most offending is concentrated in small geographic areas, and it is likely that local environmental settings (e.g., physical infrastructure) and the broader context of places (e.g., neighbourhood characteristics such as poverty and social cohesion) are predictive of the locations and timing of violence. Knowing where and when violence is most likely to happen is vital to reduce the harm it causes and ultimately to prevent it.

In the UK, there is a paucity of research studies that explore the micro level locations (spatial and temporal patterns) of violence involving young people. This project seeks to address this gap by identifying the specific times and the nature of the places where violence affecting children and young people occurs. Similarly, little is known about the contextual factors that predict these concentrations, or how they may vary by age, gender, and ethnicity.

## 1.2. Research Aim

This project aims to further our understanding of when and where young people are at risk of violence by using three to five UK police force areas as case studies. It will generate actionable insights to support efforts in preventing and responding to crimes against children and young people.

## 1.3. Research question(s)

The primary research questions are:

1) Where are the highest concentrations of crime involving children and young people?
2) When do these concentrations of crime involving children and young people occur?
3) What local contextual factors best explain the location and timing concentrations of crime involving children and young people?

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

| Question Number | Interim report | Final report |
|---|---|---|
| *1* | Full initial descriptive analysis completed. | Minor additional edits to descriptive analysis completed in the initial phase.<br>Incorporate feedback from local learning events with police forces |
| *2* | Full initial descriptive analysis completed. | Minor additional edits to descriptive analysis completed in the initial phase.<br>Incorporate feedback from local learning events with police forces |
| *3* | Datasets prepared for identify statistically significant drivers and protective factors.<br>Preparation of appropriate regression models | Complete analysis of statistically significant drivers and protective factors.<br>Full analysis of drivers and risk factors at key identified times and across different age, gender, and ethnicity groups<br>Incorporate feedback from local learning events with police forces |

### 1.4. Hypotheses

Hypothesis One [$H_1$]: Related to Research Question One: Violence will be concentrated at a small number of localised places where young people are likely to be concentrated or gather (e.g. city centres, night-time economy locations, schools and colleges, retail centres, etc). ('hot spots').

Hypothesis Two [$H_2$]: Related to Research Question Two: Violence will be concentrated at a small proportion of times ('hot times' e.g., early and late night-time economy hours, immediate periods prior to/after school/college opening/closing times).

Hypothesis Three [$H_3$]: Related to Research Question Three: Violence concentrations will be shaped by identifiable risk and protective factors linked to local context (the physical environment, neighbourhood functions, presence of crime generators and crime attractors, areas of low social cohesion and collective efficacy (characterised by high population churn

and greater social diversity) and socio-demographic characteristics associated with higher victimisation risk) which will significantly influence when and where such incidents occur.

If necessary, we will adjust the hypotheses at the interim stage, if these are not feasible or practical to test.

**Underpinning Research**

A substantial body of research has established that crime, including youth violence (i.e., violence committed against, or by, children and young people), is not random, but instead clusters in specific geographic areas and at certain times. These concentrations are commonly referred to as 'crime hot spots,' though definitions vary across studies (Eck et al., 2017; Braga et al., 2019a; Smith et al., 2023). Weisburd (2015) introduced the concept of the 'law of crime concentration', highlighting a consistent pattern across cities of varying sizes and locations: a disproportionate amount of crime occurs in a small number of places. Sherman et al. (1989) were among the first to empirically demonstrate this phenomenon, showing that over half of police service calls originated from just 3.5% of residential addresses.

Spatial analysis of crime events can be traced back to early European scholars in France (Quételet and Guerry, 1981,1831) in the 19th century, gaining momentum through the Chicago School's work on urban sociology (Shaw et al., 1929; Shaw & McKay, 1942; Park & Burgess, 1925/1967). Over time, researchers shifted focus from broad urban patterns to more granular, neighbourhood-level studies. The emergence of environmental criminology in the 1970s and 1980s further refined this approach, examining how physical and social environments shape opportunities for crime. Key theories include routine activity theory (Cohen & Felson, 1979), crime pattern theory (Brantingham & Brantingham, 1984), and rational choice perspectives (Cornish & Clarke, 1986).

Several contemporary studies can be used to further explain the spatial and social contexts that are likely to influence youth offending. These include factors such as urban mobility (Newton et al., 2015, 2021), lifestyle patterns (Hägerstrand, 1970), and the presence of crime attractors and generators (Brantingham & Brantingham, 1985). Together, these frameworks help explain how the movement of children and young people through urban environments may shape their exposure to spatial risk and influence the concentration of youth-related violence. Physical features including poor lighting and visibility (Jeffery, 1971; Newman, 1972; Clarke, 1995; Armitage & Monchuk, 2010), as well as socio-demographic indicators such as poverty and inequality (McAra & McVie, 2016) also play a role. Moreover, concepts like collective efficacy (Sampson et al., 1999), social cohesion (Hipp & Wickes, 2017), and capable guardianship (Reynald, 2016) underscore the importance of how places are managed and perceived, influencing both crime opportunities and youth decision-making.

In brief and at risk of oversimplification, we suggest there are four key contextual drivers of crime to consider in this study.

(i) The intersection of potential offenders and victims, based on routine activities, lifestyles, and the social mobility patterns of young people (specific individuals may feature in both offender and victim cohorts over time). It should be noted that the data that will be utilised for analysis in the present study does not identify where a child or young person has been both a victim *and* offender (the victim-offender overlap). Consequently, the youth sample being analysed will fall into two distinct categories: victim or offender - with a focus on the time and location of the offence.

(ii) The characteristics and context of local places which shape motivations for young people to travel to/from/through these places, formed by the function(s) of that place (for example the presence of pubs, restaurants, shops, nail bars, hairdressers, transportation facilities, and schools). These may be crime attractors (which offenders perceive as attractive targets), crime generators (large groups of people brought together creating unplanned opportunities to offend), or crime detractors (types of land use that may deter crime).

(iii) The temporal use of these different land types, partly driven by formal opening and closing times, and informal use of place (parks or bus shelters where young people may congregate during hours of darkness when no services are provided). This aspect influences the time at which distinct groups of young people congregate, as well as the levels of guardianship and protection that might be afforded at these places.

(iv) These micro locations are influenced by a broader set of neighbourhood-level contextual factors, for example socio-economic and demographic characteristics of that place, poverty levels, and community willingness to collectively protect nearby environments, and natural and human-made environmental design (e.g. levels of lighting and visibility).

**Expected magnitude of the relationship:**

(i) Spatial concentrations of violence

We anticipate that violence will concentrate spatially at different scales. At the street segment level, we predict that violence will concentrate in line with Weisburd's law of crime concentration, and approximately 50% of violence will be located within 3-5% of all street segments. At the census output area (OA) and lower super output area (LSOA) we still expect to find concentrations of violence but this to be less concentrated, in line with the 80/20 Pareto Principle (Ceccato and Newton, 2024) that suggests 80% of outcomes (violence) will occur in approximately 20% of places.

The study will identify 'alternative' hotspots of violence that are often not identified using standard hotspot analysis, adopting the Minerva Process (see New Insights section below)

developed with Essex Police (Johnson et al., 2025) to focus on areas away from high-volume hotspots (driven by high population concentrations). We will also identify hot spots where violence is 'over-represented' and where violence' bucks the trend' (see 1.5 and methods section).

We will explore spatial concentrations of violence by age of the victim and or offender where available. We recognise pre-defined Home Office (HO), Ministry of Justice (MoJ) and Department of Education (DfE) classifications for children (aged 10 – 17) and young people (aged 18 – 25). Dependent on the distribution of the data and potential emergence of meaningful groupings of age from the data, the age categories utilised in the present study may differ to the pre-defined classifications of age previously mentioned. In this instance, utilising data-driven age categories would provide valuable insights relating to distinct groupings of age and concentration of violence. This is pertinent when seeking to understand how spatial and temporal patterns of young people involved in violence may likely link to their routine activities and lifestyles – rather than pre-defined administrative classifications. We anticipate that concentrations will be higher (more violence in fewer places) for younger age categories, linked primarily to crime generators, urban mobility, lifestyle theory, and routine activities. For example, social roles and educational responsibilities may vary. Those aged 12 – 15 may have similar school times and participate in similar evening and weekend activities than those aged 16-18, or 18-25. Their peer influences and identify formations also differ, for example times and locations spent with family versus peers. Their time use priorities may change between structured routines (school, homework, extra-curricular), exam stress and late evening/night socialising, or work responsibilities. Additional external factors include legal milestones such as voting, driving, and drinking alcohol. We will also explore these variations by gender and ethnicity. However, this is dependent on sample sizes and the extent to which this data is recorded or missing in the police crime data.

(ii) Temporal concentrations of violence

We expect violence will concentrate temporally at different scales although this will not be as concentrated as the anticipated spatial concentrations. We will examine crime concentrations across the 168-hour week. We expect to find patterns consistent with the 80/20 Pareto Principle. There are limited studies here; Curiel (2021) identified that 72% of robberies occurred in 20% of the 168 weekly hours but there is no similar study for young people and violence.

Additionally, we will explore temporal concentrations by seasonality. We expect variations in offending between school day and non-school day periods (weekday and weekend, as well as school holidays and non-school holidays), and between hours of daylight. Depending on the sample size, we anticipate being able to categorise the season using two categories (Spring to Autumn equinox) or high daylight (>12 hours) and low daylight (<12 hours). Final selection

will be dependent on sample size in the police recorded crime data. We expect violence to be more temporally concentrated (across fewer hours) during low daylight hours/autumn equinox, and during school days.

We will explore temporal concentrations of violence by the age of the victim and/or offender (where available). We will use the same age classifications as proposed for spatial concentrations. We anticipate that temporal concentrations will be higher (more violence across fewer hours) for the younger age categories, primarily linked to crime generators, urban mobility, lifestyle theory, and routine activities.

We will also explore these variations by gender and ethnicity. However, this is dependent on sample sizes and the extent to which this data is recorded or missing in the police crime data.

(iii) Local contextual predictors of violence

We anticipate that there will be statistically significant drivers of violence. These include land use types (proximity to secondary schools, youth clubs, bars and clubs, fast food outlets, transport hubs); in areas with lower levels of community engagement and cohesion; places close to open parks and woodlands; in areas with higher levels of predicted footfall, and in areas with higher levels of deprivation.

We will explore potential variations in local context influences on youth violence, by age (and gender and ethnicity if sample size permits), and by time of offence (using the same classification as previously stated; see Table 3.3 for a summary of time-varying predictors). There is limited information here on the magnitude of change. Whilst we anticipate some differences here, we are cautious in predicting the spatial and temporal variations of local contextual variables and their correlations with youth violence by age, gender, and ethnicity.

**New Insights**

The primary purpose of this project is to provide descriptive insights into the spatial and temporal patterns of youth violence using regression analysis. These combined police recorded crime data with open-source data sets, the latter offering local contextual information about potential drivers of violence (protective factors and risk factors).

Whilst evidence exists on these four inter-related issues, there is currently a gap concerning the spatial concentration of young people violence and the range of interconnected potential drivers and protective factors for this. The majority of current UK evidence is based on analysis in London. Moreover, there is a need to understand better how risk/protective factors are relevant not only by place (spatial) but also the temporal nuances that influence when and where young people violence occurs, and when and where it does not.

This study will deliver robust secondary analysis and, furthermore, develop learning events with the police forces we work with in this project (see Section 2.2). The aim is to demonstrate the applied learning and actionable policy this approach offers, as well as to create interactive online mapping toolkits that partners can use to support place-based data-informed approaches to preventing violence among young people.

This will replicate the established 'Minerva Process' developed with Essex Police over a period of four years. The 'Minerva Process' aims to look beyond conventional hot spot analysis, using a place-based model (i.e., mapping police crime data to and with contextual open-source data) to establish where issues occur, their key drivers and what is 'different' in the neighbourhoods/places where issues occur. The process also incorporates temporal factors into analysis to better understand how risk shifts across key time windows. Using a case study of Violence Against Women and Girls (VAWG) this process has challenged common police assumptions about the locations/hot spots of VAWG, supported Community Safety Partners (CSPs) to work to work with local communities to develop initiatives and approaches to crime reduction in hitherto non-targeted locations where VAWG was over-represented.

In addition to the YEF Interim and Final Reports we will produce password protected interactive online toolkits and local learning events for each force area we work with. At each site, partners can use this to support place-based, data-informed approaches to violence prevention. This supports the application of learning and the development of actionable policy from this secondary study.

### 1.5. Key concepts

The key concept used in this study can be considered in relation to the specified research questions. The primary outcome is violence, which will be measured by Home Office police recording standards. Children will be defined as those aged 11-17 at the time of the offence, and young people will be defined as those aged 18-25. Crime hot spots are considered spatially (crime rate per 1000 residents, over-representation, and bucks the trend, and by counts of violence by street segment). It should be noted that street segment denominators will not be used, as high resolution ambient population/footfall data is not widely available at the street segment level. Instead, we will use theory informed independent variables including street length, number of POIs and, type of street (e.g. major/minor road) to support the regression modelling. Temporal hot spots will be explored using the 168-hour week (hourly intervals from Monday to Sunday), including school days and non-school days (such as school holidays). The autumn equinox starts on 23rd September, and the Spring Equinox Is from 20th March. Table 1.5 provides a definition of each. Measurement and datasets for each are provided in Table 3.3.

## Table 1.5 Definitions of key concepts

| Terms | Definition used |
|---|---|
| Violence | According to the Home Office Counting Rules for Recorded Crime, violence is classified under the category of "violence against the person", which includes a range of notifiable offences that involve physical force or the threat of force against an individual. This category encompasses offences such as homicide, assault with injury, assault without injury, causing death by dangerous driving, threats to kill, harassment, stalking, modern slavery, and child abuse offences, among others. These offences are recorded when, on the balance of probabilities, the circumstances as reported amount to a notifiable crime, and there is no credible evidence to the contrary. The Home Office definition is operationalised through detailed recording rules to ensure consistency across police forces in England and Wales, and forms part of the broader framework for compiling official crime statistics. |
| Young People/<br><br>Children | HO, MoJ and DfE classifications of children (aged 10 – 17) and young people (aged 18 - 25) are recognised as pre-defined age classifications. The present study will identify and utilise age categories or groupings derived from the distribution of the data to provide meaningful insights. |
| Crime hot spots (spatial) | Clusters of crime concentrated in place. This will include crime rates/1000 residents; 'overrepresentation' hotspots and 'bucks the trend' hotspots; street segment clusters, and GI* hot spots. |
| Crime hot spots (temporal) | Clusters of crime concentrated in time. This will include by the 168-hour week. Additional considerations include seasonality (spring/autumn equinoxes) and school days versus non-school days. Final considerations will include hours of daylight (plus or minus 12 hours per day). |
| Weapons enabled violence | Refers to offences involving knives or other bladed or pointed articles, including both possession and use. This includes items such as razors, |

| | |
|---|---|
| | swords, axes, bayonets, forks, needles, arrows, and broken glass bottles—whether manufactured, modified, or improvised. |

# 2. About the datasets

## 2.1. Overview of datasets used

This study will primarily use police recorded crime data on violence and young people for three to five case study police force areas. This will be for period April 2022 to March 2025. In addition, the study will draw on a range of open-source data sets including Ordinance Survey (OS) Open Roads, Points of Interest and Open Greenspace; Office for National Statistics (ONS) Census 2021; English Indices of Deprivation (IoD); the University College London (UCL) Space Syntax; and Police Powers and Procedures: Stop and Search. We will also use the in-house Nottingham Trent University (NTU) Community Engagement Area Classification.

Data will be extracted from nationally available opens source datasets to geographies consistent with the case study police force crime datasets (2022-2025). Data will be extracted at the lowest available spatial unit (easting and northing, street segment, census output area, or lower super output area) and aggregated for each spatial unit included in the regression analysis (see the methodology section).

## 2.2. Primary Data Sources

The primary data source is police-recorded crime data, using 3-5 identified police force areas. Whilst other data sources have been considered (e.g., PNC data), the use of police-recorded crime data allows for more granular spatial and temporal analyses of youth violence concentration, extending beyond an offender-only focus and enabling the provision of key insights across different geography levels (e.g., street-level, neighbourhoods, Community Safety Partnerships).

Initially, the forces will be from Essex, Leicestershire, and Nottinghamshire. The police forces included in the present study have been purposefully selected to ensure appropriate representation of different crime profiles, demographic characteristics, land use and urban/rural character. It should be noted that engagement with a wider range of police forces is ongoing and further police force area data may be included in the present study, mitigating concerns surrounding the representativeness of forces.

**Table 2.2a Primary Dataset Description - *[Police Recorded Crime Data]***

| | |
|---|---|
| **Name of dataset** | Police Recorded Crime Data *(3-5)* |
| **Data owner(s)** | Relevant Police Force Area |

| | |
|---|---|
| **Type of data** | Cross Sectional/Longitudinal (3 years) |
| **Availability of data** | Data sharing agreement and stored in Secure Environment |
| **Team member(s) who will have access** | Professor Andy Newton, Dr James Hunter, Dr Erin Gibson |
| **Population/geographic coverage or sampling frame** | National – We will utilise 3-5 police force areas once data sharing contracts have been agreed upon. |
| **Years covered or survey waves** | April 2022 to March 2025 |
| **Exclusion criteria** | *N/A* |
| **Expected population/sample size (following exclusion criteria)** | 90,000 to 100,0000<br>Estimated from Essex, Nottinghamshire, and Leicestershire (see Table 3.1) |
| **Documentation** | N/A |

### 2.3. Secondary data source(s)

Secondary data sources are specified in Table 2.2 (b to j) below.

**Table 2.2b Dataset Description - *[Ordnance Survey Points of Interest]***

| | |
|---|---|
| **Name of dataset** | Ordnance Survey Points of Interest |
| **Data owner(s)** | Ordnance Survey |
| **Type of data** | Cross Sectional |
| **Availability of data** | Jisc Educational access/license |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National<br>Accuracy - Positioned to the address location usually within the building footprint. |
| **Years covered or survey waves** | April 2022 to March 2025 - *updated quarterly* |
| **Exclusion criteria** | N/A |

| | |
|---|---|
| **Expected population/sample size (following exclusion criteria)** | The ONS has over four million records nationally. Data to be extracted relevant to each POI |
| **Documentation** | https://www.ordnancesurvey.co.uk/products/points-of-interest |

**Table 2.2c Dataset Description - *[Space Syntax]***

| | |
|---|---|
| **Name of dataset** | Space Syntax |
| **Data owner(s)** | UCL |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | April 2022 to March 2025 - *updated quarterly* |
| **Exclusion criteria** | N/A |
| **Expected population/sample size (following exclusion criteria)** | Approximately 800,000 UK street segments – to be extracted to case study police force areas |
| **Documentation** | https://spacesyntax-openmapping.netlify.app/#7/54.088/-2.043 |

**Table 2.2d Dataset Description - *[Open Greenspace]***

| | |
|---|---|
| **Name of dataset** | OS Open Greenspace |
| **Data owner(s)** | Ordnance Survey |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |

| | |
|---|---|
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | *April 2022 to March 2025 – updated quarterly* |
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | Covers over 216,000 acres of land – to be extracted to case study police force areas |
| **Documentation** | https://osdatahub.os.uk/downloads/open/OpenGreenspace |

**Table 2.2e Dataset Description - *[Spatial Signatures]***

| | |
|---|---|
| **Name of dataset** | Spatial Signatures |
| **Data owner(s)** | University of Liverpool |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | *April 2022 to March 2025 – updated quarterly* |
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | Covers over 14 million records - to be extracted to case study police force areas |
| **Documentation** | https://urbangrammarai.xyz/great-britain/ |

**Table 2.2f Dataset Description - *[MODUM]***

| | |
|---|---|
| **Name of dataset** | Classification of Multidimensional Open Data of Urban Morphology (MODUM) |
| **Data owner(s)** | University of Liverpool |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | *2016* |
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | Contains classification for all England and Wales Census OAs- to be extracted to case study police force areas |
| **Documentation** | https://data.geods.ac.uk/dataset/classification-of-multidimensional-open-data-of-urban-morphology |

**Table 2.2g Dataset Description - *[Residential Mobility Index]***

| | |
|---|---|
| **Name of dataset** | Residential Mobility Index |
| **Data owner(s)** | University of Liverpool |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | *2022* |

| | |
|---|---|
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | Contains classification for all England and Wales Census LSOAs- to be extracted to case study police force areas |
| **Documentation** | https://data.geods.ac.uk/dataset/residential-mobility-index |

**Table 2.2h Dataset Description - *[Retail Centre Boundaries and Open Indicators]***

| | |
|---|---|
| **Name of dataset** | Retail Centre Boundaries and Open Indicators |
| **Data owner(s)** | University of Liverpool |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | *2023* |
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | There are 6,423 agglomerations of retail across England and Wales- to be extracted to case study police force areas |
| **Documentation** | https://data.geods.ac.uk/dataset/retail-centre-boundaries-and-open-indicators |

**Table 2.2i Dataset Description - *[Community Engagement Area Classification]***

| | |
|---|---|
| **Name of dataset** | Community Engagement Area Classification |
| **Data owner(s)** | Nottingham Trent University |

| | |
|---|---|
| **Type of data** | Cross Sectional |
| **Availability of data** | Researcher Owned |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | *2025* |
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | Contains classification for all England and Wales Census OAs- to be extracted to case study police force areas |
| **Documentation** | https://ukdataservice.ac.uk/app/uploads/hunter2025-03-05.pdf |

**Table 2.2j Dataset Description - *[English Indices of Deprivation 2025]***

| | |
|---|---|
| **Name of dataset** | English Indices of Deprivation (IoD) |
| **Data owner(s)** | Ministry of Housing, Communities & Local Government (MHCLG) |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | 2025 |
| **Exclusion criteria** | NA |

| | |
|---|---|
| **Expected population/sample size (following exclusion criteria)** | The English Indices of Deprivation measure relative levels of deprivation in 32,844 Lower-layer Super Output Areas in England. |
| **Documentation** | English indices of deprivation 2025: statistical release - GOV.UK |

**Table 2.2k Dataset Description - *[Census 2021 – Population Estimates]***

| | |
|---|---|
| **Name of dataset** | Population Estimates for Lower Super Output Areas in England and Wales by Single Year of Age and Sex, mid-2022 to mid-2024 |
| **Data owner(s)** | Office for National Statistics |
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | Available for 2022 - 2024 |
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | Contains population estimates for all England and Wales Census LSOAs- to be extracted to case study police force areas. Population estimate data is also available at MSOA and LA level and by five-year age band. |
| **Documentation** | Population estimates - Office for National Statistics |

**Table 2.2l Dataset Description - *[Police Powers and Procedures: Stop and Search]***

| | |
|---|---|
| **Name of dataset** | **Police Powers and Procedures: Stop and Search** , England and Wales |
| **Data owner(s)** | Home Office |

| | |
|---|---|
| **Type of data** | Cross Sectional |
| **Availability of data** | Open Source |
| **Team member(s) who will have access** | All |
| **Population/geographic coverage or sampling frame** | National |
| **Years covered or survey waves** | 2022 – 2025; Updated quarterly |
| **Exclusion criteria** | NA |
| **Expected population/sample size (following exclusion criteria)** | Data is included from the 43 territorial police forces in England and Wales, as well as the British Transport Police – to be extracted to case study police force areas. |
| **Documentation** | User guide to police powers and procedures - GOV.UK ; Stop and search, arrests, and mental health detentions, March 2025 - GOV.UK |

### 2.4. Access and data protection

**Police crime data**

Access will be available based on the terms of the Data Processing Contract for each individual police force. Data will be restricted to authorised and specified staff. Data will be subject to relevant data processing legislation (*General Data Protection Regulation (GDPR) Article 6* / *GDPR Article 9* / *Data Protection Act 201).* The legal basis for sharing data will be Public Task (public interest or under official authority Article 9). The research team will complete a Data Protection Information Assessment with each case study police force area. NTU also requires a data management plan to be completed.

To ensure GDPR compliance personal data shall only be used or processed for the purposes set out in the Data Processing Contract in accordance with instructions from the Force. This includes the research institution having in place appropriate protective measures against data loss, a designated data protection officer, and ensuring those with access have undergone adequate training in the use, care, protection and handling of Personal Data and are in receipt of specific instructions in respect of the secure handling of confidential and/or sensitive information. The data will only be accessed in the approved NTU secure lab, as specified in the Data Processing Contract.

The data requested includes: Crime categories (violence against the person (crime Tree lV2, Lv3 and Lv4); Date and Time of Offence (From and To Dates, From and To Times), Location of Offence (Easting and Northing); Age of Victim/Offender (if known), Gender of Victim/Offender (if known); Ethnicity of Victim/Offender (optional); Victim/Offender Relationship (If known); MO location (public place flag) , Knife Crime/ Drugs/Alcohol flag.

Methods of data transfer, data processing roles (controllers and processors), the physical environment for analysis, training requirements and other specific requirements will be specified in the Data Sharing Contract with each force. This will also specify on how long data will be kept after the end of the project and on the data destruction procedures.

**Open-Source Data**

Data will be accessed, stored, processed, and deleted in accordance with the terms of each of the products open-source access conditions. Links to each are provided in Tables 2.2 (b to h).

## 2.5. Ethics

This evaluation requires approval from Nottingham Trent University School of Social Sciences Research Ethics Committee, a sub-committee of the NTU Research Ethics Committee. The NTU research ethics policy and procedure (last revised 1st Feb 2019) promote the highest possible standards of ethical practice in the conduct of academic research. It is committed to protecting the rights, dignity, safety, and privacy of research subjects, and to minimise risks to the health and safety of researchers.

This review considers consent forms, participant information sheets, and protocols relating to security, confidentiality, anonymisation, and data retention. This includes a review of: risk of physical, emotional or reputational harm to research participants, risk of physical or emotional harm to the researcher; arrangements for recruiting research participants, and for obtaining informed consent, including, for example, copies of participant information sheets and consent forms; arrangements for assuring the security and confidentiality of personal data; arrangements for retention, anonymisation and disposal of personal data; arrangements for debriefing research participants; arrangements for reporting and dealing with any adverse reactions to the project; and ensuring that DBS checks are carried out at the appropriate level where this is required. It will also consider steps taken to mitigate all identified risks.

The University has a Code of Practice for Research (last revised May 2021) which ensures the protection of human participants, and includes policies for personal and professional research integrity, accountability and responsibility, conflicts of interest, leadership responsibilities,

research data, research misconduct, ethical requirements, and health and safety, and reporting breaches of the code of practice.

Researchers must comply with NTU's Research Data Management Policy and submit a research data management plan that explicitly address data capture, management, integrity, confidentiality, retention, access, sharing and publication. The data plan complies with relevant legislative frameworks including data protection, intellectual property, ethics, and human rights. This proposal is compliant with GDPR and the Data Protection Act 2018. All staff at NTU are required to complete annual online training and require a mandated assessment pass to ensure GDPR compliance.

The individual records from the police crime data used in this study will contain eastings, northings and time of day of offence. This data will be processed in the secure crime lab to minimise the risk of re-identification. This will be in accordance with the data processing contracts (DPC) with each force. The crime data will be and aggregated into larger spatial units (street segment and census output area) with minimum counts (as agreed in the DPC) included in the published Minerva maps which will be password protected for force access. This further minimises risk of de-anonymisation.

# 3. Methods

## 3.1. Research design

**Table 3.1: Research design**

| | | |
|---|---|---|
| **Research design** | | Secondary analysis using descriptive and explanatory statistics to examine spatial and temporal patterns of violence and young people and identify correlated local contextual factors (risk and protective factors) |
| **Dataset(s) used** | | Police Recorded Crime Data<br>Open-Source Data (see Section 2) |
| **Population of interest** | | 11–25-year-olds who are victims or offenders of youth violence in the 3-5 Police Force Case Study Areas. |
| **Size of sample population** | | Defined as the population of young people (11-25) resident in each of the 3-5 police force areas selected for this study (see data sharing agreements)<br>Provisional Estimates of total population aged 11-25<br>Essex (297,000); Leicestershire (150,000); Nottinghamshire (157,000)<br>(note: sample will only include victims and offenders of youth violence)<br>It is estimated that 15-20% of 11–25-year-olds will be victims of violence annually (YEF 2023, ONS, 2024) |
| **Stratification variable (s) (if applicable)** | | Police Force Area (3-5 forces)<br>Analysis will be conducted across all PFAs (pooled) and stratified by each PFA |
| **Primary outcome** | **Variable** | Youth violence (police crime flagged by offender and or victim aged under 25) |
| | **Measure (instrument, scale, source)** | Measure: Count of offences<br>Scale: Ratio<br>Source: Police Recorded Crime (Police Force Area) |
| **Secondary outcome(s)** | **Variable(s)** | N/A |
| | **Measure(s) (instrument, scale, source)** | NA |
| **Main method(s) to be used or tested** | | The model will primarily use Binary Logistic Regression to consider presence/absence of crime (and other binary divides). Spatial and temporal dependencies will be considered during the analysis process (see Section 3.5: RQ3).<br>In addition, the sensitivity of the regression models (to crime counts, e.g., 20 on a street) will be tested using appropriate alternative regression techniques, dependent on distribution of data (e.g., Negative Binomial Regression, Multinomial Logistic Regression, Zero Inflated Regression; see table 3.5 for descriptions of each method). |

## 3.2. Participants

The study will analyse police recorded violence offences. Participants will be young persons aged 11-25 who are present in this dataset for violent offences that occurred in the case study

police force areas between April 1, 2025,2022 and March 31, 2026. This will be a count of offenders, victims, and unique violence events recorded. Individuals will not be identifiable from this analysis, as the names and addresses of offenders and victims are not included in the police force data request. The analysis will be conducted across 3-5 police force areas. There will be an estimated 75,000 to 100,000 sample records (see Table 3.1). Provisional force areas are Essex, Nottinghamshire, and Leicester.

### 3.3. Variables and measurement

Key measurement concepts are discussed below in Table 3.3

## Table 3.3 Measurement of key concepts

| Concept | How the concept will be measured and encoded | Variable source, definition and derivation or specification | Analysis required at interim stage to test assumptions |
|---|---|---|---|
| Violence | Violence against the person as defined by the Home Office Crime Recording Rules for frontline officers and staff | https://www.gov.uk/government/publications/counting-rules-for-recorded-crime<br>Police Recorded Crime Data for each case study force area | Explore distribution of count data to determine appropriate regression models. |
| Young People/ Children | Age of victim and or offender at the time of the crime offence in years (before next birthday).<br>Provisional category of child (11-17)<br>Provisional category of young person (18-25) | Police Recorded Crime Data for each case study force area | Checks to determine completeness of age, gender, and ethnicity records. Check the distribution of each category to determine the suitability and appropriateness of further regression analysis. |
| Crime hot spots (spatial) | Clustering of violence offences (i) volume crime rates (per 1000 residential population), (ii) over-representation of violence relative to total violent offences (using location quotients), and (iii) 'bucks the trend' locations of violence in neighbouring localities (using Moran's I | Police Recorded Crime Data and Census Population Data | Checks to verify the completeness of the time category.<br>Check the distribution against each category for suitability and determine the most |

| | | | |
|---|---|---|---|
| | analysis) and (iv) clustering by street segment based on violence counts within 200m. | | appropriate further regression analysis. |
| Crime hot spots (temporal) | Clustering of crime counts by (i) time of day segmented across the 168-hour week (hourly periods from Monday to Sunday), and (ii) weekday/weekend offences; (iii) equinox periods; and (iv) daylight hours (more than/less than 12 hours).<br>Potential clustering of crime counts related to opening/closing times of POIs will also be identified by (i) just open (an hour before opening time), (ii) open (during operational hours), (iii) just closed (an hour after closing time), and (iv) closed (during non-operational hours). | Police Recorded Crime Data using 'From Time' Opening/closing time data | |
| Weapons enabled violence | Knife-related crime (knife crime) is a crime involving an object with a blade or sharp instrument. | Police Recorded Crime Data | Check for completeness<br>Check the distribution suitability and determine the most appropriate further regression analysis. |

### Other measures

Secondary data sources used to identify explanatory variables for crime concentration were listed in Section 2, along with the sample sizes for each.

To analyse temporal variations in youth violence concentrations at potential crime generators (i.e., places creating conditions that enable or encourage criminal activity) or attractors (i.e., places well-known for providing opportunity for criminal activity), the opening and closing times of key points of interest (POIs) will be identified through the generation of 'proxy' opening/closing times using online manual searches. This will use a sample of POIs (5%) across each police force area to estimate opening and closing times for each key POI. Variations of

estimated POI opening and closing times will then be cross-compared, and variations across each police force area will be tested. The opening and closing times will be used to create a variable identifying time periods when a POI is (i) just open (an hour before opening time), (ii) open (during operational hours), (iii) just closed (an hour after closing time), and (iv) close (during non-operational hours).A discussion of the potential limitations of the sampling approach outlined can be found in Section 3.8.

### 3.4. Sample size

Our sample is defined as the population of young people (aged 11-25) residing in each of the 3-5 police force areas selected for this study, who are involved in youth violence (see data sharing agreements). Provisional estimates of total population aged 11-25 in provisionally identified forces are: Essex = 297,000; Leicestershire = 150,000; Nottinghamshire = 157,000. Note that the sample will include victims and offenders of youth violence. It is estimated that 15-20% of 11–25-year-olds will be victims of violence annually (YEF, 2023; DESW, 2024). Therefore, an indicative sample size for this study is 75,000 t0 100,000 violence offences.

### 3.5. Research methods and analysis

## Table 3.5 Summary of analytical approaches

| Analytical approach | Description | RQ addressed |
|---|---|---|
| Location quotients (LQs) | LQs measure the relative concentration of a characteristic or activity in a subarea compared to the entire area. | RQ1 |
| Moran's I | Measures spatial autocorrelation to identify how closely features are clustered on a map. | RQ1 |
| Getis-Ord (GI*) | Measures spatial autocorrelation to identify clusters of “hotspots” (high values) or “coldspots” (low values). | RQ1 |
| Generalised Gini co-efficient | Measures statistical dispersion to identify the extent of variation in concentrations. | RQ1 |
| Seasonal Autoregressive Integrated Moving Average (SARIMA) modelling | Used to identify patterns in time series data with seasonal or periodic fluctuations | RQ2 |
| Regression analysis | Explores the relationship between dependent variables (i.e., the outcome variable the analysis aims to explain or predict) and independent variables (i.e., explanatory variables potentially impacting the dependent variable). The type of regression analysis conducted depends on the nature of the dependent data (e.g., continuous, categorical, observed count), as | RQ3 |

| | | |
|---|---|---|
| | well as meeting required assumptions. More specific descriptions for the types of regression used in the present study are presented below. | |
| Binary Logistic Regression | The type of regression required when the dependent variable of interest is categorical in nature, with only two possible outcome categories. Binary Logistic Regression predicts the probability of an event (i.e., one of the two outcome categories) occurring. | RQ3 |
| Multinomial Logistic Regression | The type of regression required when the dependent variable of interest is categorical with two or more unordered outcomes. Multinomial logistic regression predicts the probabilities of the different possible outcomes. | RQ3 |
| Poisson Regression | The type of regression required when the dependent variable of interest consists of count data. It should be noted that two key assumptions of Poisson regression are that the count data follows a Poisson distribution and that the mean and the variance of the model are equal. | RQ3 |
| Negative Binomial Regression | The type of regression required when the dependent variable consists of count data but violates the key assumption that the mean and the variance of the model are equal. In other words, when there is more variation (i.e., overdispersion) or less variation (i.e., underdispersion) than expected in the data. | RQ3 |
| Zero-Inflated Regression | The type of regression required when there are more zeros in the dependent count variable than expected under standard count distributions (i.e., Poisson or negative binomial). Zero-inflated models are typically used where the data contains two distinct types of zero: structural zeros (i.e., the outcome will always be zero), or sampling zeros (i.e., as part of standard count distribution). | RQ3 |

***RQ1:*** Drawing on the scale of police-recorded violence at the Census Output Area level, we will initially produce three measures of crime concentrations: volume crime rates (per 1000 residential population), over-representation of violence relative to total offences/violent offences (using location quotients), and 'bucks the trend' locations of violence in neighbouring localities (using Moran's I analysis).

Following this, violence will be snapped to street segments (based on the OS Open-Roads database) using GIS software to produce frequency maps of street segment violence

concentrations. We will explore the impact of using 50-250m buffers to snap violence incidents to the nearest street segment having previously used 100m buffers which identified 97% of VAWG incidents in public places.

We will identify hot spots of violence using non-administrative boundaries (i.e., 100-250m grids – sensitivity to be tested) to identify statistically significant hot spots of young people violence using Getis-Ord (GI*) analysis.

We will explore the extent to which these differing spatial concentrations vary using the Generalised Gini co-efficient.

Depending on the violence sample size, we will also replicate this analysis by victim characteristics (age profile e.g. 11-17, 18-25), gender (male, women, and girls) and ethnicity.

**RQ2:** We will use SARIMA models to examine variations in violence at different geographic locations, utilising relevant time intervals, including: the 168-hour cycle, comparisons between daytime and weekend periods, school opening and closing hours, seasonal variations and finally, comparisons between opening and closing times of key points of interest.

Given the concerns over autocorrelation in temporal data, SARIMA is the method adopted and accepted by multiple researchers who examined the temporal impact of the COVID-19 Pandemic that suppressed the levels of public place crime offences.

Depending on the violence sample size, we will also replicate this analysis by victim characteristics (age profile e.g. 11-17, 18-25), gender (male, women and girls) and ethnicity.

**RQ3:** Having identified the spatial and temporal concentration of violence, we will use Binary Logistic Regression (running different versions of binary divide, including presence/absence, in top 10%/not in top 10%) and multivariate analysis to identify a range of theoretically informed potential drivers and protective factors for violence primarily extracted from open-source data. Our previous research suggests that this is the most appropriate method. However, we will test the sensitivity of the models (to crime count, e.g., 20 on a street) using appropriate alternative regression methods (e.g., Poisson/Negative Binomial Regression), based on the distribution of the data.

These contextual data sources include ONS Points of Interest data (to identify potential crime generators and crime attractors), relevant 2021 Census variables (to establish the socio-demographic context of different locations including age, gender and ethnicity), English Indices of Deprivation (2025),the UCL space syntax database (to measure pedestrian movement and 'eyes on the street'), NTU's community engagement area classification (to measure social cohesion and collective efficacy levels), and open parks and green spaces. A full list was provided in Section 2 of this plan.

Having identified the significant drivers and protective factors for violence, we will conduct temporal analysis to explore (a) the extent to which these micro-pockets of crime are driven by temporal time of day characteristics; and (b) the impact of these factors in shaping patterns of violence at different times of the day and days of the week.

Spatial dependencies will be accounted for when conducting regression analysis by incorporating spatial weighting and spatial lag terms into the regression models. Spatial independent variables will also be added to the regression models (e.g., length of street, street type and the number of POIs on a street). Temporal dependencies will be accounted for by the inclusion of time-window independent variables to the regression models. The time windows will be selected using the Minerva Approach and will reflect population and crime 'rhythm', underpinned by routine activities and lifestyle theory. In addition, time window variables will be used to explore the role of POIs as crime generators or attractors. While POIs are always present and so have constant influence, this is not stable 24-7, particularly when closed or not in use.

To control for error related to multiple testing, Family Wise Error Rate (FWER) and False Discovery Rate (FDR) corrections will be applied. These are likely to be the Holm-Bonferroni method (for FWER) and the Benjamini-Hochberg procedure (for FDR), but alternative correction methods will be considered if more appropriate.

**Modelling uncertainty and conducting inference**

We will initially assess our data (see Table 3.3) to explore which regression methods are most appropriate for our analysis. This is expected to primarily be Binary Logistic Regression, although alternative forms will be considered (e.g., Multinomial Logistic, Negative Binomial, or Zero Inflated Regression). Table 3.5 summarises each of the regression methods mentioned. Our statistical regression models will include appropriate confidence intervals, identifying coefficients, p-values, and odds ratios. Where appropriate we will identify model fit (AIC, BIC, log likelihood).

## 3.6. Missing data

The following data fields in the police recorded crime may be incomplete (time field, ethnicity field, gender field, age field).

The extent to which we can compare the magnitude of change between age, gender, and ethnicity will depend on the completeness of these records.

Where the time from is incomplete, we will explore inferring times of violence offences from the time reported and/or time to fields.

The biggest concern here is ethnicity field. As police forces are required to return notifiable offence counts to the Home Office, and this is assessed as part of the HMFRICS review we anticipate other fields will be complete (>95%). This reflects our experience of working with previous data from police forces.

### 3.7. Additional analysis and robustness checks

Diagnostics to assess for appropriate regression modelling: mean, variance, linearity, multi-collinearity, normality of residuals, model fit, dispersion tests (for over-dispersion and zero inflation), residual normality, model fit (AIC/BIC).

As the data covers a three-year period, we will use ANOVA, Kruskal-Wallis, or Kolmogorov-Smirnov Tests (depending on the distribution of the data) to determine if the police crime data is comparable or statistically different over each of the three years. Previous experience suggest that the data is likely to be comparable. If significant differences are identified, we will explore spatial and temporal concentrations of crime across each year and model these against the open-source contextual data.

### 3.8. Other potential limitations or sources of bias

Police-recorded crime data has several limitations and sources of bias that can affect its reliability. One concern is underreporting, as noted in the Crime Survey for England and Wales, where many crimes go unreported and are therefore not recorded. However, for micro-level analysis, no alternative data source exists that facilitates our proposed regression methods. We acknowledge the availability of the R package *rcme* that tests levels of measurement error in crime data where crime rate is an independent or dependent variable in the model. However, the present study does not create crime rates for regression modelling. Instead, population will be utilised as an independent variable and presence/absence of crime (as a binary variable) or crime count will be utilised as the dependent variable. Consequently, use of the *rcme* package is not appropriate for the present study.

Additionally, variations in recording practices across police forces and changes in classification rules can lead to inconsistencies over time and between regions. We have chosen a time with no major changes in recording standards for violence. Administrative errors, such as misclassification or duplication, may cause bias. We will verify data quality by testing for completeness and reviewing HMRCFRS data quality reports for the identified three-year period for each selected force. Biases also arise from selection effects, where only detected or reported crimes are included, and surveillance bias, where areas with more policing may appear to have higher crime rates regardless of actual prevalence. Operational priorities may skew the data toward certain crime types, and we will use the published stop and search data

to test for this. These factors can be compounded by policy-induced changes, such as shifts in legislation or reporting incentives, which may artificially inflate or suppress crime statistics. We will examine national interventions during the specified period. We have avoided national clear build host rollouts, as well as the COVID period.

To ensure our research appropriately considers race equity our proposed spatial and temporal analysis will:

- Analyse the nature of young people violence incidents by age, gender, and ethnicity to identify distinctive patterns across different victim cohorts.
- Breakdown the identification of volume, over-representation, and 'bucks the trends' crime concentrations by age, gender, and ethnicity to establish whether different victim cohort hot spots occur within the same or different localities.
- Undertake different versions of the spatial/temporal modelling of potential drivers and protective factors relating to young people violence incidents by victim cohort to identify whether these operate universally by age, gender, and ethnicity – or vary in significance or magnitude across different victim cohort dimensions.

The limitations of the sampling approach for POI opening and closing times should be noted. The outlined approach produces a representative sample of opening and closing times for POIs in the police force case study areas. It is recognised that opening and closing times may significantly vary within and between police force areas and as such, findings related to POI opening/closing times may not be generalisable to wider areas. For example, areas where cultural factors or population density may impact operating hours (e.g., rural areas) or where seasonal factors may impact operating hours (e.g., coastal or tourism-driven areas). Consequently, future studies should replicate the sampling approach outlined to identify variations in opening and closing times for areas of interest. In addition, premises may also close earlier than formally published closing times if they are quiet. This is recognised as a potential limitation – but without granular data on footfall the opening/closing times are used as a proxy.

The research team will explore whether local authorities hold opening/closing times and or footfall for the police force areas used in the study – although some forces may span multiple local authorities. If necessary, we will adjust the methodology at the interim stage, for example increasing sampling to 10% for 1-2 forces, should the proposed sampling strategy be impractical to implement.

# 4. Project management

## 4.1. Risks and mitigations

Table 4.1 below identified potential risks for this project and mitigation strategies.

### Table 4.1 Risks and mitigations

| Number | Risk | Likelihood (*Low/Medium/ High)* | Mitigation |
|---|---|---|---|
| 1 | Fail to agree data processing contract for a minimum of three police forces within timeframe | Medium | Data is being requested from six to ten sites<br>We have had early communication with ten police force areas to secure appropriate access at a minimum of three sites.<br>We have established good working relationships with several forces from previous research using similar data, and our legal teams are familiar with police DPIAs. This enables easier engagement and the continuous development of working relationship with police forces. |
| 2 | Data requires secure environment | Low | NTU has a secure crime lab that meets industry standards in terms of storing and analysing this data (on non-networked PCs). Access to the lab, PCs, and individual folders is restricted as required. |
| 3 | Low sample size to detect significant correlations between violence data and local context data | Low | We have worked with similar data and are confident a three-year period will be sufficient to identify relationships at each police force area. We can pool data across sites to increase sample sizes. |
| 4 | Changes to the research team | Low | We can draw on the wider NTU governance, whereby alternative staff with relevant skills can be utilised as a |

| | | | |
|---|---|---|---|
| | | | contingency. We have a pool of temporary 'UniTemps' staff we can use to support the project on a short-term basis (1-12 weeks) should we need to replace any staff. |
| 5 | Information & data loss | Low | Our procedures comply with ISO/IEC 17799 for Data Security and the requirements of Hannigan report on Data Handling Procedures (2008), GDPR and the UK data protection Act (2018); NTU also have a secure crime lab and Newton and Thompson have 'Approved Researcher' status with the UK Data Service (UKDS). |
| 6 | Delivery within budget and time | Low | The PI and Co-Is have extensive experience of project management. Internally we will use MS Teams to ensure consistent communication. We will agree monthly meetings with the core research team. We will ensure regular progress reporting using YEF templates and monthly meetings with YEF (as agreed). |
| 7 | Data is not quality assured | Low | The NTU research data management plan ensures, at the outset, that we include plans for the storage, transfer, and retrieval of data, including appropriate backups, during the live phase of the project. It also includes a data archiving policy that will meet with the explicit requirements of the YEF. We will obtain NTU ethical consent consistent with the ESRC Framework for Research Ethics. |

### 4.2. Timeline

A project timeline is provided below:

## Table 4.2 Timeline

| Date | Activity | Staff responsible/leading |
|---|---|---|
| Months 1-3 | A1: Develop Data Sharing Agreements with 3-5 Police Forces (DPIA agreements) for six years of police recorded crime data. | Andy Newton |
| Months 1-3 | A2: Seek a favourable ethical opinion from NTU's Schools of Business, Law, and Social Sciences. | James Hunter/Rich Pickford |
| Months 1-3 | A3. Recruitment of a Research Fellow (Erin Gibson) – skills include understanding of police crime data, and data science (proficient in r or python), and understanding of GIS. | Andy Newton/Becky Thompson |
| Months 1-3 | **A4: Produce Statistical Plan using YEF existing templates (First Draft 13th October 2025)** | Andy Newton/James Hunter |
| Months 4-9 | A5: Secure data transfer of police crime data to NTU's secure crime lab, with data cleaning and matching by key geographies (street segments, 250m grids, and census output areas). We will create data scripts in R/Python to enable replication of this process at each of the police forces. | Andy Newton/James Hunter |
| Months 4-9 | A6: Capture of open-source data on local and neighbourhood context. | Andy Newton/James Hunter |
| Months 4-9 | A7: Capturing of school, young people clubs and other key POI opening and closing times, using a sample approach. | Andy Newton/Erin Gibson |
| Months 4-9 | A8: Descriptive Spatial Analysis of locations of violence hot spots. | Andy Newton/ Erin Gibson |
| Months 4-9 | A9: Descriptive Temporal Analysis of when violence hot spots occur. | Andy Newton/Erin Gibson |

| | | |
|---|---|---|
| Months 10-15 | A10: To identify statistically significant drivers and protective factors, development of regression models at three spatial scales – what are risk and protective factors? | Andy Newton/ in Gibson |
| Months 10-15 | A11: To assess the drivers and risk factors at key identified times (during and after school closure, weekday and weekend, and school holiday versus non-school holiday). Run regression models to compare time periods. | Andy Newton/ Erin Gibson |
| Months 10-15 | **A12: Interim Report (First Draft 4th September 2026)** | Andy Newton/ Erin Gibson |
| Months 16-18 | A13: Learning Events/Dissemination with Local Partners (from police force areas who supplied project data) | Rich Pickford/Andy Newton |
| Months 16-18 | A14: Creation of bespoke project website to provide wide access to project findings. | Andy Newton/James Hunter |
| Months 16-18 | **A15: Production of the Final Report (First Draft 4th December 2026)** | Andy Newton/ Erin Gibson |
| Months 16-18 | A16: Publication of Journal Papers – subject to YEF and police force agreement. | Andy Newton |
| Months 16-18 | A17: Explore Cross Force Learning Event (seek funding through e.g., the Society for Evidence-Based Policing or the College of Policing, or host online). | Rich Pickford/Andy Newton |

## 5. References


Armitage, R., & Monchuk, L. (2010). Sustaining the crime reduction impact of designing out crime: Re-evaluating the Secured by Design scheme 10 years on. *Security Journal*, 24(4), 320–343. https://doi.org/10.1057/sj.2010.6

Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in Crime*. New York: Macmillan.

Ceccato, V. & Newton, A., 2024. Systems Thinking for Sustainable Crime Prevention: Planning for High-Risk Places. Routledge. ISBN 9781032249858

Clarke, R. V. (1995). Situational Crime Prevention. Crime and Justice, 19, 91–150. https://doi.org/10.1086/449230

Cohen, L., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. American Sociological Review 44(4), 588–608. https://doi.org/10.2307/2094589

Cornish, D. B., & Clarke, R. V. G. (1986). *The reasoning criminal: Rational Choice Perspectives on Offending*. Springer Verlag.

Hägerstrand, T. (1970). What about people in Regional Science? *Papers in Regional Science*, 24(1), 6–21. https://doi.org/10.1007/bf01936872

Hipp, J. R., & Wickes, R. (2017). Violence in urban neighborhoods: A longitudinal study of collective efficacy and violent crime. *Journal of Quantitative Criminology*, 33(4), 783–808. https://doi.org/10.1007/s10940-016-9311-z

Jeffery, C. R. (1971). *Crime Prevention through Environmental Design*. Beverly Hills, CA: Sage.

Johnson, A., Ellis, C. Hunter, J., Newton, A., Pickford, R., and Hill, R. (June 2025). The Minerva Approach Blueprint. Taking a Partnership Approach to Hot-Spot Mapping and Understanding Community Based Issues. Final Report Submitted to the Eastern Region Innovation Network (ERIN)

McAra, L., & McVie, S. (2016). Understanding youth violence: The mediating effects of gender, poverty and vulnerability. *Journal of Criminal Justice*, 45, 71–77. https://doi.org/10.1016/j.jcrimjus.2016.02.011

Newman, O. (1972). *Defensible Space*: Crime Prevention through Urban Design. New York: Macmillan.

Newton, A. D., & Felson, M. (2015). Editorial: crime patterns in time and space: the dynamics of crime opportunities in urban areas. *Crime Science*, 4(1). https://doi.org/10.1186/s40163-015-0025-6

Newton, A. D., Felson, M., & Bannister, J. (2021). Introduction to the special issue: Urban Mobility and Crime Patterns. *European Journal on Criminal Policy and Research*, 27(3), 307–311. https://doi.org/10.1007/s10610-021-09501-7

Office for National Statistics. (2024). The nature of violent crime in England and Wales: Year ending March 2024.

https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/thenatureofviolentcrimeinenglandandwales/yearendingmarch2024

Park, R. E., Burgess, E. W., & McKenzie, R. D. (1967 [1925]). *The City. Suggestions for Investigation of Human Behavior in the Urban Environment*. University of Chicago Press.

Quételet, A. (1984 [1831]). Research on the propensity for crime at different ages. Translated by S. F. Sylvester. Cincinnati: Anderson.

Reynald, D. M. (2016). *Guarding against crime. Measuring Guardianship within Routine Activity Theory*. Routledge. London. https://doi.org/10.4324/9781315586007

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighbourhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328), 918–924

Shaw, C. R., Zorbaugh, F. M., McKay, H. D., & Cottrell, L. S. (1929). *Delinquency areas: A Study of the Geographic Distribution of School Truants, Juvenile Delinquents, and Adult Offenders in Chicago.* Chicago: University of Chicago Press.

Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas*. Chicago: University of Chicago Press.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place. *Criminology*, 27(1), 27–56. https://doi.org/10.1111/j.1745-9125.1989.tb00862.x

Weisburd, D. (2015). The Law of Crime Concentration and the Criminology of Place. *Criminology,* 53(2), 133–157. https://doi.org/10.1111/1745-9125.12070

Wilson, J. Q., & Kelling, G. (1982). Broken windows: The police and neighbourhood safety. *Atlantic Monthly* 249(3), 29–38.

Youth Endowment Fund. (2023). Children, violence and vulnerability 2023. https://youthendowmentfund.org.uk/reports/children-violence-and-vulnerability-2023/summary/