Youth Endowment Fund
Administrative Data Guidance

PERU Policy Evaluation & Research Unit

YOUTH ENDOWMENT FUND

Manchester Metropolitan University

Youth Endowment Fund (YEF) Administrative Data Guidance

Using administrative data for youth crime and violence outcomes in evaluations

August 2024

Mark Ellison

Will Cook

PERU Policy Evaluation & Research Unit

Manchester Metropolitan University

Administrative Data Guidance

Table of Contents

About the Youth Endowment Fund

The Youth Endowment Fund (YEF) is a charity with a mission that matters. We exist to prevent children and young people becoming involved in violence. We do this by finding out what works and building a movement to put this knowledge into practice.

Children and young people at risk of becoming involved in violence deserve services that give them the best chance of a positive future. To make sure that happens, we'll fund promising projects and then use the very best evaluation to find out what works. Just as we benefit from robust trials in medicine, young people deserve support grounded in the evidence. We'll build that knowledge through our various grant rounds and funding activity.

Just as important is understanding children and young people's lives. Through our Youth Advisory Board and national network of peer researchers, we'll ensure they influence our work, and we understand and are addressing their needs. But none of this will make a difference if all we do is produce reports that stay on a shelf.

Together, we need to look at the evidence, agree what works and then build a movement to make sure that young people get the very best support possible. Our strategy sets out how we'll do this. At its heart, it says that we will fund good work, find what works and work for change. You can read it here.

For more information about the YEF or this report, please contact:

Youth Endowment Fund
C/O Impetus
10 Queen Street Place
London
EC4R 1AG
www.youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

Registered Charity Number: 1185413

About the Policy Evaluation and Research Unit (PERU), Manchester Metropolitan University

Established in 2007, the Policy Evaluation and Research Unit at Manchester Metropolitan University is a multi-disciplinary team of evaluators, economists, sociologists, and criminologists. We specialise in evaluating policies, programmes and projects and advising national and local policy-makers on the development of evidence-informed policy. We work in the UK and Europe for clients and funders including UK government departments, local government, the voluntary sector, and the European Commission. What makes our work distinct is our emphasis on methodological rigour, our knowledge of multiple methods, and our broad expertise across different sectors.

Mark Ellison (Research Fellow) m.ellison@mmu.ac.uk

Dr Will Cook (Reader) w.cook@mmu.ac.uk

Administrative Data Guidance

1 Introduction

1.1 Purpose and overview

At the heart of the Youth Endowment Fund (YEF) approach to evaluation is the use of rigorous research methods, such as randomised controlled trials (RCTs) or quasi experimental designs (QEDs) to find out whether an intervention, project or activity is effective. Effectiveness can be measured in many ways and depends on what the intervention aims to change - the outcome. YEF uses the follow data sources to measure the effectiveness of the projects it funds:

1) **Measurement of self-reported outcomes *within* the evaluation period** – Because we want to prevent children and young people from becoming involved in violence and crime in the first place, we fund many interventions, projects or activities that support children and young 'upstream' on involvement of crime or violence. That means that we focus on projects that aim to change outcomes (or risk and protective factors) that are related to violent and criminal behaviour later.
2) **Measurement of outcomes administrative data *within* the evaluation period** - Ultimately, YEF's mission is to build the evidence base for what works in reducing crime and violence. Therefore, wherever feasible, evaluators are encouraged to select a crime and violence outcome as the evaluation's primary outcome wherever possible.
3) **Tracking the long-term outcomes of projects *after* an evaluation has finished** – YEF's data archive involves collecting, storing, and archiving data on participants so they can be followed-up and their outcomes assessed against criminal justice records in future years.

YEF has guidance on 1) and 3), but no guidance around 2). This report is designed to fill that gap. YEF's outcomes framework and measurement review provides comprehensive guidance on measuring risk and protective factors (1), with additional guidance on the core measures used in many YEF evaluations: the Strengths and Difficulties Questionnaire (SDQ) and the Self-Reported Delinquency Scale. However, YEF will always want to measure crime and violence directly through administrative data wherever possible. This is facilitated by our data archive which enables researchers to access data on YEF funded trials. YEF has provided detailed guidance for evaluators on the data archive.

There are two limitations on relying on the data archive as the only source of access to administrative data on crime and violence. First, the evaluation must have finished, the report published before the data is archived. Second, an approved researcher must have applied to access the YEF data via the ONS secure research service (SRS)[1]. This builds in a considerable time lag before we can draw conclusions about a project's effectiveness in reducing crime. Therefore, YEF always want evaluators to access administrative data with crime and violence records within the evaluation period wherever possible.

This report outlines the administrative data that is available that is likely to be of use to those conducting evaluations of Youth Endowment Fund (YEF) funded interventions. The purpose of the document is to inform evaluators of the key strengths and weaknesses of such data and how to approach arranging access to the data to support evaluation. We hope it will be useful to evaluators carrying out YEF evaluations that have crime or violence outcomes as a primary of secondary outcome, as well as other researchers wishing to make use of this administrative data.

What are administrative datasets?

*"Administrative data are a **by-product of administrative systems developed primarily for operational purposes**. Administrative data are used extensively in the **compilation of many sets of official statistics** about a wide range of topics." (Office for Statistical Regulation, 2024)*[2]

Examples of administrative data include:

- **Local Police data** which includes police recorded crime data collected by one of the 43 local police forces across England and Wales. Local Police data includes details of crime events (i.e., offence type, location, date/time) or suspects / offenders (age, gender, ethnicity)).
- **Police National Computer (PNC) data** is a national dataset which includes information about police cautions and court convictions held on individual offenders in England and Wales. The Ministry of Justice (MOJ) receive a data extract to examine offenders' convictions over time to conduct reoffending analysis by offender characteristics.
- **Hospital Episode data** which includes Accident and Emergency (A&E) attendance or hospital admission for injuries associated with violence. Data on individual (patient) episodes include demographics (age, gender, ethnicity).

[1] https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice soon to be replaced by the Integrated Data Service (IDS) https://integrateddataservice.gov.uk/about-the-integrated-data-service

[2] https://osr.statisticsauthority.gov.uk/guidance/administrative-data-and-official-statistics/quality-assurance-of-administrative-data-case-examples/administrative-data-part-1/

- **Linked datasets** across the criminal justice system and other government datasets (for example Ministry of Justice (MOJ) Data First) which enables accredited researchers, across government and academia, to access anonymised, research-ready datasets ethically and responsibly. Data First aims to unlock the potential of the wealth of data already created by MoJ.

Administrative data has a number of advantages over other data that maybe collected. Compared to sample surveys, administrative data has a much larger sample size and thus leads to increased power in evaluations. This not only means that confidence in estimates of intervention effects can be more easily obtained, but also that the estimation of effects for sub-groups of the population and for rarer crime outcomes are more feasible. In addition, data collected on offending outcomes as part of the operation of law enforcement and the criminal justice system may be more likely to be reliable than self-reported behaviour from individuals, particularly those who may face incentives to under report their offending (e.g., those on licence).

There are however drawbacks as well; for individuals to appear within administrative dataset, they are likely to have been criminalised (especially for PNC). Offenders must have met a threshold in terms of offending severity, frequency or age of criminal responsibility for prosecution. Therefore, survey-based measures may be more adept at capturing more sensitive or refined measures of offending / offending behaviours, and in some cases may be more likely to record offending behaviour than official sources (see Basto-Pereira & Farrington, 2019[3] and Thornberry & Krohn, 2000[4] for discussion). Administrative data is usually restricted in terms of the depth of the variables that are collected and by its nature may not contain the detail necessary to measure the intended effects of an intervention, particularly effects that may form part of mechanisms of change. It may also be biased by the focus of law enforcement efforts at a particular point in time towards particular crime groups and/or socio-demographic groups.

In addition, as administrative data is rarely collected for the purposes of research and evaluation, there may be poorly defined and understood procedures for researchers to obtain data access.

[3] Basto-Pereira, M., & Farrington, D. (2019). Lifelong conviction pathways and self-reported offending: Towards a deeper comprehension of criminal career development. British Journal of Criminology, 1-18.
[4] Thornberry, T.P., & Krohn, M.D. (2000). The self-report method for measuring delinquency and crime. Measurement and Analysis of Crime and Justice, 4, 33-83.

1.2 Envisaged use

1.2.1 Individual level administrative data

In studies that assess the effect of interventions that aim to reduce the propensity for individuals to offend or re-offend, individual level data is required on offence outcomes. In most cases this will be in circumstances where individuals (and parent / guardian) consent for their personal records can be accessed, which may be challenging in some contexts. YEF provide guidance on this in the 'Data Protection information for YEF evaluations' report[5] .

It is important to recognise that such data will naturally be an inaccurate record of an individual's actual offending behaviour, as administrative data typically only records offending that requires some contact with the police and/or the Criminal Justice Service (CJS).

Individual level administrative data can be provided as an identifiable dataset (i.e., unique reference number (URN), name, date of birth, address) or pseudo anonymised[6], which does not allow the individual to be directly identified.

In many YEF evaluations, police and other criminal justice datasets are used as a secondary outcome measure alongside primary self-reported measures including the: Self Report Delinquency Measure (SRDM) (Smith & McVie, 2003), Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) or Warwick-Edinburgh Mental Well-being scale (WEMWBS) (Tennant *et al*., 2007). Individuals may receive different levels of activity (dosage) on their intervention. Therefore, it is important to link datasets which represent the various inputs, outputs, and outcomes (i.e., operational intervention data, self-reported measures, and police / criminal justice administrative data) collected around an individual. To enable this to happen, it is important to receive a linking variable such as identifiable data field (e.g., a unique reference number (URN)) or if not available, a prior pseudo anonymised reference to enable data matching.

1.2.2 Geographical Area / Place-Based level administrative data

In other cases where interventions seek to reduce the incidence of crime and / or antisocial behaviour, rather than necessarily crime committed by certain individuals, area-based aggregates of criminality / offending are usually required. For the purposes of evaluation, these data (e.g., local police recorded crime data or nationally available data from data.police.uk or office for national statistics (ONS)) can

[5] https://youthendowmentfund.org.uk/wp-content/uploads/2021/07/YEF-Data-Guidance-Projects-and-Evaluators.pdf

[6] 'Pseudonymisation' of data (defined in Article 4(5) GDPR) means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. / Pseudonymisation refers to techniques that replace, remove, or transform information that identifies individuals, and keep that information separate. (ICO, 2024)

either be accessed as administrative or census area level data or, if bespoke geographies or offence types are required, these can be generated/requested by aggregating up offence and recorded crime data from precise geo-locations into the desired geography.

Geographical level data, like individual level data, do not present a complete picture of crime in an area. This is because large proportions of crime are not either reported to or recorded by the police. The Crime Survey of England and Wales (CSEW) identified that in 2020 only 42% of comparable crime incidents are reported to the police (see ONS, 2022[7]). Concerns about crime recording by the police (HMICFRS, 2018[8]), resulted in police recorded crime no longer classified as a 'national statistic designation' in 2014 (Office for Statistical Regulation, 2023[9]). This was due to concerns about the quality and consistency of police crime recorded practices, with variation between different forces (HMICFRS, 2018). HMICFRS has undertaken a rolling programme of crime data integrity inspections at a police force level, to understand the levels of under-recording[10]. Health datasets (e.g., A&E attendance) are now regularly being used to examine violence trends under the Information Sharing to Tackle Violence (ISTV)[11] initiative.

1.3 Information flows through the criminal justice system and data access points

Figure 1 illustrates a high-level overview of the information flows within the criminal justice system (police and courts / probation) and the various *data access points* (through the Local Police Forces, the Ministry of Justice (MOJ), the Home Office (HO), the Single on-line service (So-lS), individual Health Trusts or data Linkage (Data First) for research and evaluation purposes (these are represented as circles). Criminal Justice administrative datasets start with police operational processes (which are presented within the blue box). A call for service on the police (incident) is logged on the local police incident recording system. These incidents may result in a crime and an arrest. Information is recorded onto a local police crime recording system (i.e., details about the crime and details on the individual(s) associated with a crime event). After an offence is recorded by the police, a suspect may be identified, and an arrest might be made. **When a suspect is arrested, police must also enter their details onto the PNC system as quickly as possible**[12]. If a suspect is charged, they will progress through criminal justice system including courts, probation, and prison. These organisations collect their own

[7] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmarch2022
[8] https://hmicfrs.justiceinspectorates.gov.uk/our-work/article/crime-data-integrity/
[9] https://osr.statisticsauthority.gov.uk/publication/systemic-review-outline-police-recorded-crime-statistics-quality-review/#:~:text=Police%20recorded%20crime%20statistics%20for,of%20police%20crime%20recording%20practices.
[10] https://hmicfrs.justiceinspectorates.gov.uk/our-work/article/crime-data-integrity/crime-data-integrity-programme-judgment-criteria/
[11] https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1594-information-sharing-to-tackle-violence-minimum-dataset
[12] PNC Recording process https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/873384/PNC_v5.0_EXT_clean.pdf

administrate data from operational processes and systems (i.e., the suspect is charged and prosecuted, the case goes to trial and if found guilty the offender is sentenced) and the current disposal of the offender is logged. Some of these details are also recorded on the Police National Computer (PNC).

Data standards for police operational information and data entities, developed by Home Office / National Police Chiefs Council (NPCC) are used to support the consistent and accurate recording of data across the 43 territorial police forces in England and Wales. The (Person, Object, Location, Event) POLE standards[13] are constructed from a combination of 'data components' and 'validation rules'. They describe people, objects, and locations associated with events. However, in this guidance we will use interchangeable terms which relate people (individuals / perpetrators / offenders / nominals[14]), events (incidents, crimes, episodes) and locations (points, addresses and areas – e.g., Census, electoral or administrative aggregated level data).

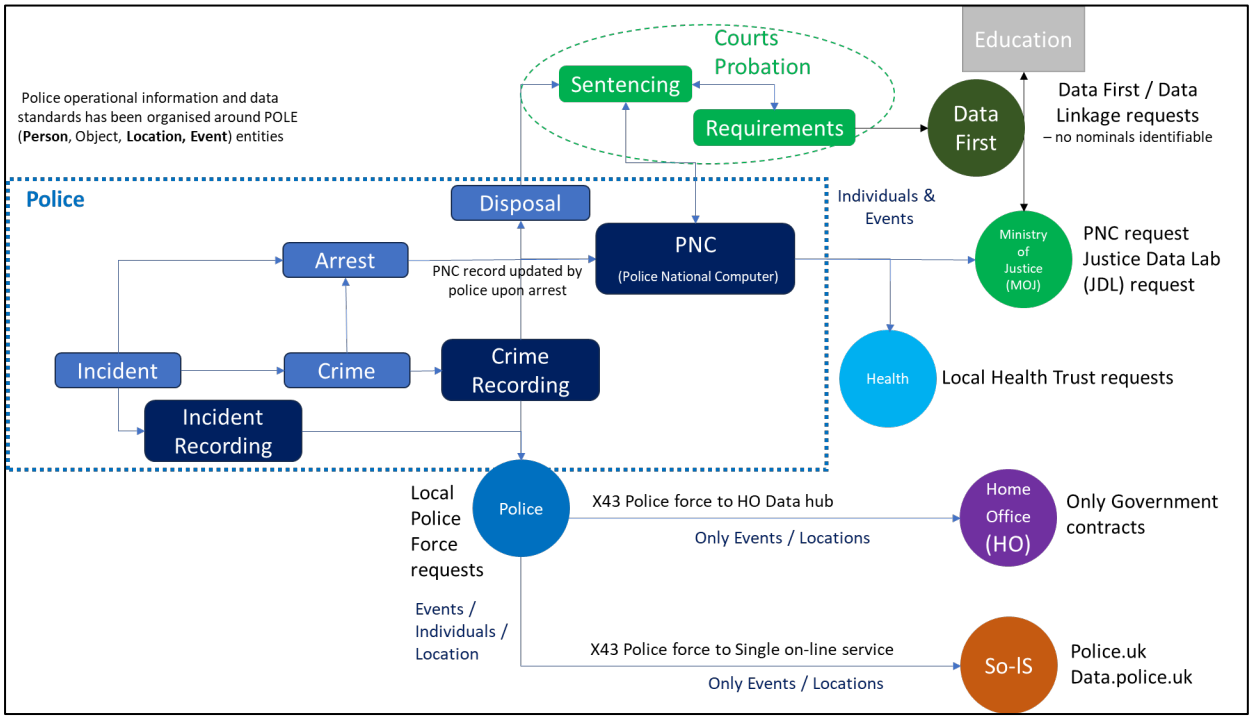

Figure 1: Information flows and access points

[13] https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/disclosure-logs/dei-coordination-committee/2023/274-2023-pole-data-standards-catalogue-v1.1-1-1.pdf

[14] "Individuals (nominals) who have come to the notice of police as offenders, suspected offenders or whose details have been recorded for another policing purpose" (CoP, 2023), these could be a could be a victim or witness or not related to offending behaviour, such as missing people, licencing or road traffic collisions.

1.4 Structure of guidance document

Section 2 provides an overview of data access processes, including common steps and processes required to develop an Information Sharing Agreement (ISA) and best practice in operational data management for evaluation.

Section 3 maps the administrative datasets that are likely to be of use for YEF evaluation; these include Police National Computer (PNC), local police data, hospital episode data, MOJ Data First datasets and other relevant datasets. This section will provide a description of the dataset, which variables are important for evaluation and the key considerations when using these data.

Section 4 provides individual data access procedures for the key datasets. This section will also include case studies of YEF (and other) evaluations, illustrating innovation and best-practice in data access for evaluation.

Figure 2 illustrates a flow chart of the possible trial designs focused on Individuals, places, and cohorts. For each design there are a number of datasets which could be utilised. This flow chart provides signposting for sections of this guidance report.

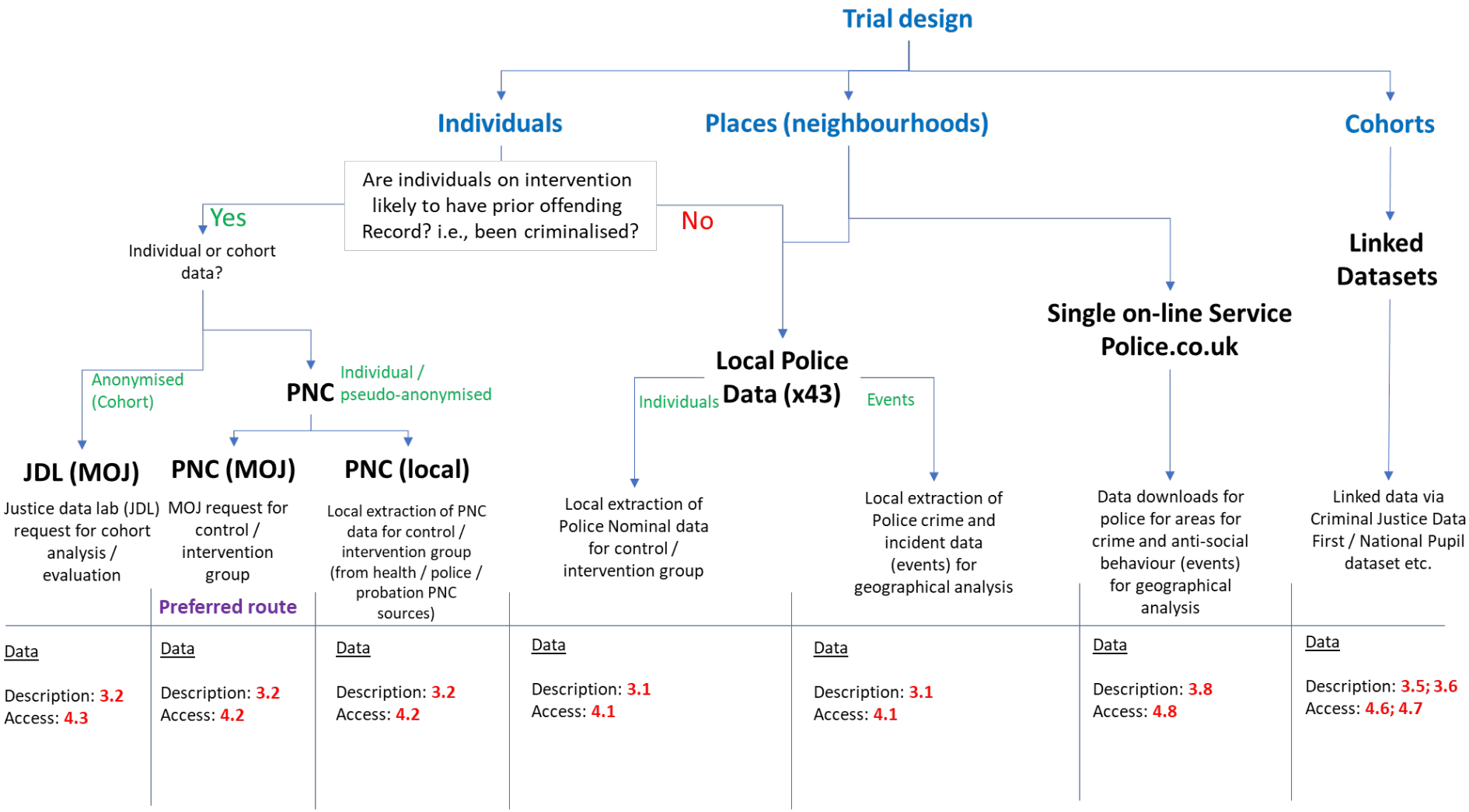

Figure 2: Flow chart of datasets by Trial design

Note: Data section provide signposting to data description (section 3) and how to access these data (section 4)

2 Overview of data access processes

2.1 General process

This section provides an overview of common steps and processes for data access. This section is primarily modelled on accessing Local Police data; however, these steps are also relevant to other datasets. Figure 3A illustrates the stages pre-data access to the point of Information Sharing Agreement (ISA) sign-off.

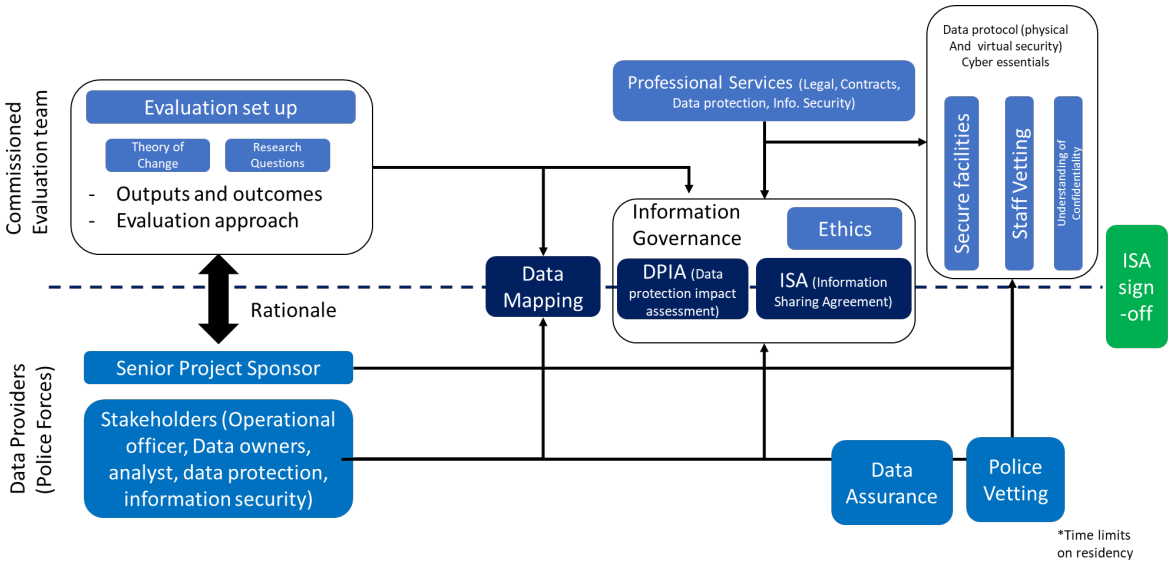

Figure 3A: Pre-data access processes

Figure 3B illustrates the operational data management processes, and best practice to ensure data are available of appropriate quality for evaluation purposes.

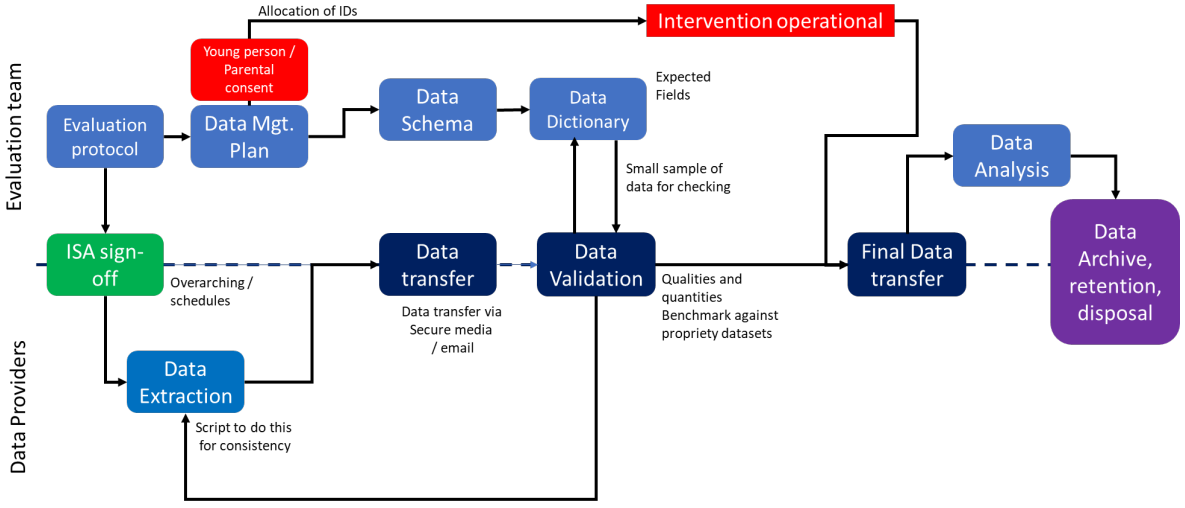

Figure 3B: Operational data management processes

These processes are examined in more detail below.

2.2 Identifying a point of contact

It is important to identify a point of contact to access data for evaluation purposes. This will vary by dataset. This section covers local police datasets in detail, offering best practice, where there are multiple routes for identifying a point of contact and stakeholders, and a section on more direct routes for Police National Computer data from MOJ.

2.2.1 Local Police Datasets

There are numerous challenges in accessing **local police data from individual police forces**. The crucial factor to successfully achieving this is to identify key stakeholders within in the Police Force(s) an evaluation team is working with.

Firstly, it is always good practise to identify a **senior project sponsor,** ideally someone who is at a more senior rank including senior officers for example Chief Officer ranks, Superintendents or Chief Inspectors. Preferably, this is someone who works at Police Headquarters rather than local police stations. Contacting either an External Relations, Performance, Policy, or Research team is a good start. However, this will vary by police force and prior / existing relationships[15], establishing these may be challenging and requires persistence. Collaborations including the N8 Policing research partnership (N8PRP, 2024)[16], Society of Evidence Based Policing (SEBP)[17], together with College of Policing[18] are a useful starting point to identify appropriate contacts. Moreover, contacting offices of Police and Crime Commissioners (PCC), Community Safety Partnerships (CSPs) and local organisations delivering YEF interventions may also support in brokering appropriate contacts.

Alongside a senior project sponsor there will be several stakeholders to engage with. These may include Operational officers who may be involved with the intervention. Police staff who act as data owners, for example these may be from an analytical team including data analysts / researchers. In addition to this there is likely to be engagement with staff from Data protection and Information Security. It is a good idea to make inquiries, identify key stakeholders and map out key personnel in these roles. The police force may also have a role (e.g., Strategy & Policy Officer - External Relations & Performance) or body that acts as a research or academic liaison role.

One of the major issues working with police stakeholders is first their initial identification, in many forces, officers and police staff change role on a very frequent basis (6-12 months). Therefore, over the duration of an evaluation project, researchers may work with different staff in the same role. It is

[15] from consultation with evaluators, many research teams already have established links with individuals in police forces to support the brokering and access to data. It is noted that it is challenging to establish these relationships with appropriate individuals when starting a project, however this is an essential stage for evaluation teams to work through.
[16] https://www.n8prp.org.uk/home/about/
[17] https://www.sebp.police.uk/
[18] https://www.college.police.uk/research/support-research

good practice to have senior project sponsor buy-in and keep formal correspondence and documentation. New staff performing these roles, may be more cautious and risk averse, and will need briefing appropriately, and a trusting relationship to develop. Therefore, it is important to develop a clear project and evaluation brief which can be used to communicate with these stakeholders.

2.2.2 Other data routes

For accessing other datasets, the same processes and practice are seen as best practice, however with a different set of point of contacts and stakeholders. For accessing the **Police National Computer (PNC) data**, the main point of contact is the data linking team (datalinkingteam@justice.gov.uk). On initial application, you will work with an identified point of contact at MOJ, who will provide advice and support to refine your application, so it meets the needs of your evaluation. Likewise, with specific local health datasets or Data First datasets there are specific specialist point of contacts. More details are in section 4 of this guidance.

2.3 Data Mapping

For any evaluation and project set up it is vitally important to develop a theory of change (TOC)[19] and research questions. The TOC will contain inputs, activities, outputs, and outcomes of the intervention under the broader evaluation approach. It is important that these elements are incorporated into the data mapping processes with administrative datasets.

For this phase of work, it is strongly recommended that evaluators work alongside organisational staff (e.g., a police officer or analyst) with knowledge of the local police system and data. Each of the 43 police forces are operationally independent and have their own crime recording system (however, some forces use the same software supplier). They can advise on the precise data fields required, resources required and the estimates of time to extract these data. It is important to factor in any resource required by the organisation(s) providing information into the evaluation project plan for timescales (lead-times) and budgeting estimates – they may require additional funding.

This data mapping exercise will support the drafting of the information sharing agreements (ISA) and data schedules and inform, any data protection Privacy Impact Assessment (DPIA) or ethical requirements.

2.4 Developing the Information Sharing Agreement (ISA)

At this point, the commissioned evaluation team's professional services (for example legal, contracts, data protection, and / or information security etc.) will be required to advise and support the research

[19] A description and illustration of how and why a desired change is expected to come about, as a result of activities and inputs.

team and liaise with the police force data providers relevant team's, data protection, information security etc., to draft, formalise and finalise any data sharing agreements, to enable the exchange of data. There are various guidance documents from local police forces, College of Policing[20], Home Office[21], MOJ[22] and NHS[23] on how to do this. YEF have a series of comprehensive guidance on the Evaluation data archive for evaluators https://youthendowmentfund.org.uk/evaluation-data-archive/

The requirement of providing data may rest upon, the approval of secure facilities, secure data transfer processes, staff vetting. Named contacts (researchers, organisation IT support, etc.) may need to be vetted, and approved staff may need to sign (individually) an Understanding of Confidentiality document (which forms part of the data schedule under the ISA).

2.5 Vetting

It is important to engage with the staff (researcher / support) vetting early in the project. Evaluation staff may require non-police personnel level 2 (NPPV2)[24] or level 3 (NPPV3[25]). Please be prepared to provide a range of personal information (including details about family, any siblings and their relationship, cohabitants, financial information, social media handles etc.) This information may take time to collate, so engage early and be prepared for lead times. The vetting process may take several months due to demand and staffing in police vetting teams. Also, for certain researchers there may be issues with issues with obtaining vetting, due to UK residency limits. NPPV requires at least three years residency, prior to vetting taking place by a police force.

2.6 Information sharing agreements (ISA)

Establishing an Information Sharing Agreement (ISA) is a legal requirement for data access. This section only provides a brief overview. More details can be found on specific information sharing pages of police force and government websites. For examples, The College of Policing website[26] provides a very detailed overview. Individual organisation, usually have their own templates, which are then completed jointly, and terms negotiated between the two parties. Moreover, ISAs usually have an overarching element (tier 1) and individual schedules (tier 2) which are either project specific or cover different data items.

[20] https://www.college.police.uk/app/information-management/information-sharing
[21] https://assets.publishing.service.gov.uk/media/652cefa56b6fbf000db7567a/data-sharing-guidance-criminal-justice-system.pdf
[22] https://assets.publishing.service.gov.uk/media/62038afa8fa8f510b357cc44/data-sharing-guidance-researchers.pdf
[23] https://www.england.nhs.uk/wp-content/uploads/2022/06/B0989-NHS-violence-prevention-and-reduction-standard-guidance-notes.pdf
[24] NPPV level 2: (full) unsupervised access – police material/information up to OFFICIAL-SENSITIVE with occasional access to SECRET
[25] NPPV level 3: unsupervised access – police material/information up to SECRET and occasional access to TOP SECRET
[26] https://www.college.police.uk/app/information-management/information-sharing

Tier 1 – The Overarching Data Processing Agreement

1. The Parties
2. Purpose
3. Definitions
4. Uses, Disclosure and Publication
5. Data Protection and Subject Rights
6. Freedom of Information
7. Security
8. Review, Retention and Disposal of Data
9. Confidentiality
10. Audit
11. Review of Data Processing Agreement
12. Training
13. Complaints and Breaches
14. Disputes
15. Term, Termination and Variation
16. Indemnity
17. Signatures

A. Understanding of Confidentiality

Tier 2 Individual schedules

To supplement an overarching information sharing agreement, individual schedules which are specific to individual projects or datasets which require particular data processing arrangements also need to be established.

2.7 Data Protection Privacy Impact Statement (DPIA)

The requirement to conduct a Data Protection Impact Assessment (DPIA) is set out in the data protection legislation[27]. This may be embedded within institutions Ethical Requirements for Research. A DPIA will address the nature, scope, context and purpose(s) behind the collection and use of personal information. Importantly, it helps researchers / institutions to consider any risk to individuals that is associated with the data processing and how to mitigate those risks. It is important that risks should be considered in terms of likelihood and severity of any impact on the individuals. More details on DPIAs are available on the YEF data archive. It may be appropriate and more efficient to work with data providers / data controllers to develop a shared DPIA which is subsequently utilised by each organisation.

[27] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/accountability-and-governance/data-protection-impact-assessments/#:~:text=You%20must%20do%20a%20DPIA%20before%20you%20begin%20any%20type,or%20serious%20impact%20on%20individuals.

2.8 Secure facilities and data transfer

A requirement prior to data transfer (and is some cases a condition of an ISA being agreed) is the establishment of secure facilities within an organisation receiving data, i.e., a trusted research requirement. To handle sensitive and personal information or provide technical products and services, an organisation will require Cyber Essentials Certification.[28] Individual organisations may also require external network penetration testing[29] to comply with ISO (International Organization for Standardisation) 27001, the UK Data Protection Act 2018 and the UK GDPR.

For the secure transfer of information, Criminal Justice Secure eMail (CJSM)[30] or Egress Switch[31] can be used to transfer data between people working in criminal justice public, private and voluntary organisations. The Criminal Justice Secure eMail (CJSM) permits the transfer of information up to an equivalent of 'OFFICIAL', including 'OFFICIAL SENSITIVE', in a secure way. Egress is used to share sensitive information by a number of UK Councils, Government Departments and NHS and other healthcare organisations.

In other cases, volumes of data may be larger (the current CJSM limit is 20MB) and alternative data transfer methods are required including. The preferred method is via Secure USB; however, this may require travel to host organisations (e.g., Police HQs; Darlington for PNC). Other approaches evaluators have used include:

1) Digital airlocks which are fully audited by a Safe Haven[32] team.

2) Individual participant files transferred through a whitelist website – drop into airlock into virtual computer two factor authentication (2FA). A trusted Research Environment model for which ISO accreditation is usually required.

It should also be noted that PNC data accessed via MOJ are not permitted to be hosted in cloud-based environment / systems.

2.9 Operational data management processes

The Operational data management process (see figure 2B) are important steps for evaluation teams to take and have been identified through best practice (see section 4).

[28] https://www.ncsc.gov.uk/cyberessentials
[29] https://www.ncsc.gov.uk/guidance/penetration-testing
[30] https://www.cjsm.net/
[31] https://www.egress.com/

[32] https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens

YEF evaluations require an evaluation protocol[33]. It is also good practice to develop a data management plan on how data are to be collected, processed, validated for the purpose of the evaluation. This will include:

- **Gaining consent and identifying the legal process for processing data** – In order to obtain ethical approval for a project and to comply with accepted ethical standards for research, researchers will generally need to obtain the informed consent of individual participants for their involvement in the research. GDPR recital 33 notes that research must act in a manner that is 'in keeping with recognized ethical standards for scientific research', and ethical review boards will usually expect informed consent (though not always). This is distinct from the legal basis for processing data. For example, a person may be asked to consent to participate in research (ethical basis) and told that, if they agree to participate, data about them will be processed for a task in the public interest (legal basis). Here, the legal basis for data processing will be 'public task' rather than consent. Further information on this important distinction is in YEF's Data Protection Guidance for Evaluators
- **Privacy notices and consent forms** – Once an appropriate institutional ethical review process have been followed and a legal basis for data processing has been established consent forms and privacy notices need to be developed at the start of an intervention and a requirement for data sharing with police forces, health trusts and MOJ for PNC access. Depending on the individuals within the trial, this may require informed consent and / consent of parents / guardian (See YEF Guidance). It is not possible (very challenging) to retrospectively gain consent from young people / guardians for their data to be shared for evaluation purposes. Capturing consent may be challenging and there may be the need to oversample to ensure sufficient numbers evaluation, depending on the type and scope of evaluation. However, it must be noted that for some YEF evaluations this may not be possible due to a small number of participants (from a small number of corresponding schools). It may also be appropriate to undertake assessments of the cognitive ability for young people to engage with an intervention and understand expectations within data collection.
- **Testing data transfers and validation** – including undertaking 'dummy runs' to test processes. Validation of intervention data against propriety organisation performance reporting – e.g., counts and trends over a period of time, to ensure data are consistent.

[33] https://youthendowmentfund.org.uk/wp-content/uploads/2022/03/17.-YEF-evaluation-guidance-March-2022.pdf

- **Data schemas / data dictionary** – understanding both the quantities and qualities of the data – making individual assessments of the completeness of data, where there are limitations and how this may impact of sample sizes, statistical power.
- **Data retention / deletion** - aligned to Information sharing agreement and retention schedules.

More details on the YEF evaluation and data arching process are available in the Data archive and privacy statement document.

3 Description of available datasets

3.1 Local police data

3.1.1 Description

Local police data is collected for operational purposes by each of the 43 territorial police forces across England and Wales. Data are collected on a range of IT systems including Incident Management / Command and Control, Crime Management, Custody and Case Management. Collectively, these are referred as local police data (LPD) in this guidance report. Local police data comprises the two main data collections covered in figure 1.

- Local police data on incidents (calls for service) that covers all demands placed upon the police requiring assistance, including crime, anti-social behaviour, public safety, and road traffic incidents. Some of these incidents may result in one or more crimes. These feed into the local police recorded crime datasets.
- Local police data on crime events and associated nominals (suspects and offenders) is held in crime recording systems. These are used for operational policing purposes. Some of these data are transferred onto the Police National Computer (PNC) (an overview of the national PNC is included in the subsequent section 3.2). Local police records are not a wholly separate data entity to the PNC. The key distinction between locally held police data and the PNC data is that i) access to what is effectively PNC data can be arranged via local police forces rather than attempting to directly access the PNC, and ii) there are instances where the local police record will be more complete than the corresponding record on the PNC. However, **local police data only covers a specific police force's territorial geography**, and only contains crime events for that area and offenders who have committed offences there, whereas PNC covers England and Wales. Therefore, data would need to be collected from multiple police forces for interventions which span multiple areas.

In most cases an evaluation utilising local police data will consider one of two dimensions:

- The evaluation is **focused on individuals** receiving an intervention to interrupt or prevent offending. [**Individual]**
- The impact on an intervention is in a defined **geographical area [Place Based Analysis]**

Evaluations based on Individuals.

In most cases, for an individual based evaluation, it is important to identify participants who are on an intervention programme, and for these to be subsequently identified and linked in a police data set. Individuals in a police data set are generally classed as nominals (and may appear as victims, suspects, offenders, or witnesses within a police data set). However, for some evaluations, police data may be used locally to identify the cohort, the evaluation team undertake randomisation into intervention and control groups prior to intervention approaches, therefore participants are not aware they are part of a trial.

To enable the linking of data it is important to have key variables including forename, surname, date of birth, address & post code, gender and if available a PNC number (typically referred to as PNC ID). PNC numbers can generally be collected from criminal justice organisations (with appropriate information sharing agreements in place) on an individual or cohort level.

The request that evaluators will need to make of the police include the need to identify individuals on interventions (control and intervention groups) as nominals within the police data. Evaluators need to ensure that the police extract any antecedents (i.e., previous contacts or events), which includes arrests and offending behaviour as either a *suspect* or *offender*, and any corresponding disposals which are flagged on the system.

Local Police Datasets (LPD) are extracted from one of the 43 territorial police force's crime recording systems. Police crime recording systems are relational databases which aim to capture the *what, where, when, who* and *how* of a crime. Systems are constructed around key tables including; crime events (what), offences (what), time of offence (when), nominals (who), location gazetteer (where), and crime outcomes. They contain unique identifiers for a crime event and nominals, which allow data associated with crimes and individuals to be linked. Other data collected includes; i) the *modus operandi (MO)*, a description of how the crime took place (i.e., the methods and means), and, ii) associated flags to provide details around types of offences and circumstances (domestic abuse, knife crime, hate crime, repeat victim, etc.).

3.1.2 Available variables

Data Identification

Typical data fields required for this type of evaluation would include:

- a **unique reference number (URN)**, **the date of offence** [committed / reported],
- **the offence type** (which would adopt the Home Office notifiable offences and counting rules[34]), and
- **the crime outcome type** (which is linked to the Home Office Crime Outcome Framework[35]).

It is worth noting that there may be a lag in the data on the crime system before a suspect is identified and a decision made on their disposal. Also, due to the 'operational nature' of these data, individuals may appear as *suspects* in the dataset, and then change to an *offender* once a charge / summons is indicated as their disposal.

Type of Offence – This code will link to the to the Home Office (HO) notifiable offences. There are 1,600+[36] notifiable offences, these are aggregated into groups of offences (Crime Tree). All groups are divided between 'Victim based crime' and 'Other crimes against society'.

- Victim based crime includes Violence against the person, Sexual offences, Robbery, Theft offences and Criminal damage and Arson offences.
- Other crimes against society includes Drug offences, Possession of weapon offences, Public order offences, and Miscellaneous crimes against society. These are further sub-divided into smaller groups.

Please note that there are differences in offence codes between the HO (notifiable offences) and MOJ for court proceedings. A code lookup up classification document is available and used to define the Home Office offence codes used in the court proceedings database (these underpin the data used in the Criminal Justice Quarterly statistics publication).

The Home Office offence codes can also be linked to the Cambridge Crime Harm Index[37] or ONS Crime Severity Score[38] to estimate levels of severity for each crime, based upon the *harm* caused to victims. The index or score uses prior sentencing data to calculate the typical number of days a convicted offender would spend in prison for each individual type of offence. For example, the ONS Severity

[34] Home Office (2023) Home Office Crime Recording Rules for frontline officers & staff
https://www.gov.uk/government/publications/counting-rules-for-recorded-crime
[35] Home Office (2021) Crime outcomes in England and Wales: Technical Annex
https://www.gov.uk/government/statistics/crime-outcomes-in-england-and-wales-2020-to-2021/crime-outcomes-in-england-and-wales-technical-annex
[36] https://www.gov.uk/government/publications/counting-rules-for-recorded-crime
[37] https://www.crim.cam.ac.uk/research/thecambridgecrimeharmindex
[38] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeseverityscoreexperimentalstatistics

Score uses the average number of days in prison, therefore, Homicide would have a severity weighting of 7,832 per crime; Wounding (2,088 per crime); Assault without Injury (13 per crime) etc. These are particularly useful for looking at changes in the severity or seriousness of offending, in quasi-experimental evaluation designs, when comparing individuals and cohorts during for a period prior to and after and after intervention.

Crime Outcome (Disposal Type) - Crime outcomes are assigned to each police recorded crime. These include charged/summonsed, out-of-court orders (both formal and informal), crimes taken into consideration (TIC), where the prosecution is prevented on not in the public interest, if there are any evidential issues, or where the police have stated that their investigation is completed, and no suspect have been identified. Only a small proportion of police recorded crimes (approx. 10%) have suspects identified and assigned as charged / summonsed, TIC or have been given out-of-court disposals (Home Office, 2023). In about 40% of crimes the outcome is assigned as Investigation complete - no suspect identified (Home Office, 2023).

The table below illustrates some of the key variables in a nominal dataset.

Table 3.2.1: Key variables in a Nominal Dataset

Variable	Description	Type	Notes
Unique Reference Number	Unique identifier for each individual		
Date Reported	Date Crime reported to police	Date	There are usually multiple dates in LPD; date reported, date recorded, earliest and latest date committed
Type of Offence	Home Office Offence Type	String (code)	This code will link to the to the HO notifiable offences. There are 1,642+ offences, these are aggregated into groups of offences (Crime Tree)
Crime Outcome type	Outcomes assigned to offences	Numeric (code)	This code will link to the HO Crime Outcome Framework and includes Charged/Summonsed, out-of-court orders, etc.
Date of Birth	Date of Birth of suspect	Date	For calculating age of individual
Gender	Gender of suspect	String	
Ethnicity	Ethnicity of Suspect		16+1 / 18+1[39] (which may be self- defined or observed)
Status	Suspect / Offender		These are changed retrospectively if a suspect has been identified, and has been convicted at court

[39] HM Government (2018) Criminal Justice System Exchange Data Standards Catalogue Notification of Change: Introduction of 'Self Defined Ethnicity – 18+1' Standard https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/691544/self-defined-ethnicity-18plus1.pdf

Geographical based evaluations

For geographical based evaluations, it is important to clearly identify a geographical extent using administrative boundaries (e.g., ward, Lower Layer Super Output area (LSOA) or an intervention boundary (defined by the intervention e.g., area or grids). These can be both an intervention and control area(s). There are different types of evaluation design for place-based interventions. There are strong evaluation designs which include randomisation and weaker quasi-experimental methods. Types of design are listed below (based on the Magenta Book (HM Treasury, 2020: p13-24), with descriptions by Smith *et al*., 2023 p36-37) apply geographically, but can also be used at an individual / cohort level as well.

Methods involving randomisation include:

- **Cluster-randomised trials** - Areas (or groups) are allocated in a random way to intervention or control.
- **Stepped-wedge design** - If all areas eventually will get the intervention, but not at the same time, e.g., because of resource constraints, it is possible to randomize for a place in the queue.

Quasi-experimental methods include:

- **Interrupted time series** - Time-series data are utilised to estimate trend and to describe what happens when the trend is "interrupted by" an intervention.
- **Difference-in-difference** – These builds on interrupted time series. By estimating trends in control areas, it is possible to strengthen the inference by comparing differences before and after an intervention period.
- **Regression discontinuity design** - Sometimes a cut-off threshold is introduced e.g., to restrict access to a programme offered to people, groups, or areas. Those just above and just under the threshold are probably very similar in all other respects (except being offered the PBA or not). Comparing their results offer an estimate of impact.
- **Use of concurrent control areas with pre- and post-measurements** - Helps in contrasting findings.
- **Propensity score matching** - A selection algorithm is used to improve the selection of control areas instead of using a "manual" procedure (exact matching). Data from the intervention sites and comparison sites are combined, the probability of being selected as an intervention site is estimated (called propensity scores) and those scores can be used in matching.

- **Synthetic control methods** - A pool of potential comparable observations, using historical data, is used to model how areas would have fared without the intervention. Divergence between the actual observations and the "synthetic" control gives the impact estimate.

More details on place-based evaluation designs are available in Smith et al., (2023) Evaluating Place-Based Approaches: a review of methods used

Data Identification

Typical data fields required for geographical based evaluations would include geographical / spatial dimensions, the date of offence, the type of offence (crime code from Home Office Counting rules plus 'crime tree' aggregation). The National Crime Recording Standard (NCRS) sets the expectations for crime recording within the law, regarding timeliness of recording, victim-focused and consistency across police forces.

Geographical dimensions – These are associated with the location of the crime event. This may be an actual location or an area. Police forces typically use a local land and property gazetteer (LLPG)[40] for addresses (buildings) and other locations (for example points which represent streets or parks etc.). Data within local police data typically include the x-y coordinates (British National Grid (BNG)), the address, as well as both Census geographies[41] (for example Census Wards, Lower Layer Super Output Areas (LLSOAs), Output Areas (OAs) and other administrative geographies (police beat / wards). These data can be examined in a geographical information system (GIS) to understand crime patterns and visualised as thematic or heatmaps identifying hotspots.

Table 3.2.2: Key variables in a Crime Event Dataset

Variable	Description	Type	Notes
Crime Reference Number	Unique identifier for each crime event		
Date Reported	Date Crime reported to police	Date	There are usually multiple dates in LPD; date reported, date recorded, earliest and latest date committed
Offence	Home Office Offence Type	String (code)	This code will link to the to the HO notifiable offences. There are 1,642+ offences, these are aggregated into groups of offences (Crime Tree)
x-coordinate	Eastings coordinate – location of crime	Numeric	British National Grid (BNG)
y-coordinate	Northings coordinate – location of crime	Numeric	British National Grid (BNG)
Address	Address of the location of crime (inc. postcode for buildings) or location details (e.g., street / park etc.)	String	Address and location qualities may vary between police forces.
Special Interest Markers	A marker to indicate a type of crime (e.g., domestic abuse / knife crime)	String	Markers are used to flag types of crime which cannot be identified through the home

[40] GeoPlace what is a LLPG? https://www.geoplace.co.uk/local-authority-resources/guidance-for-custodians/how-to/about-the-role/what-is-an-llpg
[41] ONS (2021) Census 2021 Geographies https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies

			office offence types. These are typically used for crime including knife crime and domestic abuse which span multiple home office offence types. Individual police force systems may have on coding systems

Incidents / Calls for Service

A call for service on the police (also known as incidents) are reported by the public (999/111), other emergency services, or observed by the police. The police assess each call for service and a decision on the threat, risk and harm posed by the situation, which informs the deployment of resources to deal with the incident. This determines the response grade, and the length of time police are expected to attend the incident, i.e., an emergency response should be within 15 minutes. The recording of data is governed by the National Standard for Incident Recording (NSIR)[42]. These data include calls regarding crime related incidents, and non-crime related incidents including anti-social behaviour (ASB), public safety and welfare (e.g., domestic incidents, mental ill health, vulnerable persons, missing persons etc.) and road traffic incidents, together with police administration and qualifiers (e.g., calls made with good intent by the public, however no perpetrator present when police arrive). Individual types of incidents can be identified by a series of closing codes and qualifiers. See NSIR guidance for more details. These systems also contain a range of semi-structured / unstructured data, which are focused on details of the initial call for service, how the incident was responded to, and ongoing operational activities by the police and partner agencies.

Table 3.2.3: Key variables in an Incident Dataset

Variable	Description	Type	Notes
Incident Reference Number	Unique identifier for each incident		
Date Reported	Date Crime reported to police	Date	Date and time when incident was logged
Incident Type	Broad grouping of incident type	String (code)	These are usually individual crime types, ASB (anti-social behaviour), PSW (public safety and welfare), Transport, or internal administration or qualifiers
Opening codes	Codes which determine different types of incidents (at point of incident being logged)	String (code)	Aligned to the groups identified in the NSIR (see link above) for example domestic dispute, mental ill health etc.

[42] National Standard for Incident Recording (NSIR)
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/116658/count-nsir11.pdf

x-coordinate	Eastings coordinate – location of Incident	Numeric	British National Grid (BNG)
y-coordinate	Northings coordinate – location of Incident	Numeric	British National Grid (BNG)
Address	Address of the location of Incident	String	
Closing codes	Codes which determine different types of incidents (once the incident has been resolved)	String (code)	Aligned to the groups identified in the NSIR (see link above) for example domestic dispute, mental ill health etc.

3.1.3 Key considerations

Recorded crime data may be most appropriate for those evaluations looking to estimate models where the intervention is geographically based, exploiting the data's geographic identifiers and coverage of crime for which no offender is identified.

- Home Office recorded crime data does not cover all offences. They only include those that are deemed 'notifiable offences' (see here for list) and will not include most summary offences; those tried in a magistrates court or by the police issuing a Penalty Notice for Disorder or a Fixed Penalty Notice , e.g. motoring offences, TV licensing, disorder etc.
- The recorded crime data does not measure the severity of an offence.
- As noted earlier in section 1.2.2, recorded crime will not capture all crime as some crime will remain unreported.
- The completeness of the data is dependent on data being received from Police Forces; not all police forces may have up to date data at a given point in time.
- There may also be issues with nominals in local police data. There may be duplicates (the same individual with multiple nominal records. Therefore, a verification or matching exercise would be required to confirm. Also, the PNC number may be used and an additional unique identifier. However, this is usually undertaken periodically by the police under the Code of Practice on Police Information and Records Management (PIRM, 2023[43]) which replaced the Management of Police Information (MOPI, 2005)).

[43] https://www.college.police.uk/guidance/police-information-and-records-management-code-practice

3.2 Police National Computer (PNC) data

3.2.1 Description

The Police National Computer (PNC) "is a large administrative database containing information about police cautions and court convictions held on individual offenders in England and Wales. The PNC is regularly updated as new information about particular individuals becomes available". (MOJ 2022)[44]

PNC data is collected by police forces and operationally used for law enforcement, other policing, and safeguarding responsibilities. Therefore, the data is reviewed updated for accuracy and currency. offending outcomes are collected for individuals from the age of ten.

3.2.2 Available variables

The PNC is a collection of several databases (including persons, property, vehicles etc.). The Ministry of Justice (MOJ) receive an extract of the PNC. The extract focusses on individuals cautioned or convicted since 2000 and where applicable their offending history. The PNC focusses on recordable offences, the offenders convicted or cautioned for them, and the outcomes received by these offenders. Recordable offences are defined as offences that can attract a custodial sentence plus some additional offences defined in legislation. Some non-recordable offences are also included on the PNC, particularly when they accompany recordable offences in the same case. The main difference between PNC data and the information from other sources, such as court data, is that the PNC does not include a range of less serious summary offences (such as TV licence evasion and a range of motoring offences). Variables include limited personal characteristics of the individual offender and details of the offence as well as disposal details. An excerpt of variables of likely use to evaluators is set out in the tables below. Where, table 3.1.1 include a list of variables required to request an extract from MOJ PNC. Table 3.1.2 incudes a list of variables commonly extracted from PNC for analysis and evaluation.

Table 3.1.1: Data variables required to undertake a match for PNC extract

Variable	Description	Justification
URN	Unique local identifier	necessary to link data back to original data for supports direction of support and local outcomes specific to the participant.
Forename	Individual's Forename	Necessary to link the data (due to potential absence of PNC ID). Personal details are common between datasets and are

[44] MOJ (2022) The Data First Project: An Introductory User Guide
https://assets.publishing.service.gov.uk/media/62149d4ed3bf7f4f0655016c/data-first-user-guide-version-7.0.pdf

		necessary to allow data match to PNC.
Surname	Individual's Surname	"
Date of Birth	Individual's Date of Birth	"
Gender	Individuals gender (Males, Females, Unknown)	"
Postcode	Individual's postcode	"
PNC ID (if available)	PNC identifier	Necessary to link to PNC data (if available)

Table 3.1.2: Data variables typically provided in a PNC extract for evaluation purposes

Variable	Description	Justification
URN	Unique local identifier	Necessary to link data back to original data for supports direction of support and local outcomes specific to the participant.
Case type	Court / Out of Court Disposal	Necessary to identify the range of disposals given, particularly those resulting in custodial periods
Court or Caution date	Court or Caution date	Allows offences to be identified relative to the date of referral to the intervention (i.e., Intention to Treat (ITT))
Offence ID	Number of Offences in Incident	Allows multiple offences to be grouped into distinct incidents
Home Office Offence Code		Identifies the type of offending
Home Office Offence Category	Home Office Offence Category	Identifies the type of offending
Disposal Category	Disposal category for each recorded disposal	Allows disposal history to be identified
Disposal Date	Disposal date for each recorded disposal	Allows disposal history to be identified
Disposal Duration	Disposal duration for each recorded disposal	Allows disposal history to be identified
Disposal Amount	Amount for first Disposal (for fines) (for each recorded disposal where relevant)	Allows disposal history to be identified
Adjudication code	Guilty / Not Guilty	Notes: only guilty verdicts were will be required
Primary Offence	Was this the primary offence? (Yes / No)	Where multiple offences are involved, this identifies which offence is the primary offence for sentencing purposes
Long Offence Description		Enable re-categorisation of offence codes
Disposal Rank	Ranking of the disposal, in terms of severity, compared to other disposals for that offence	Helps to distinguish between disposals in terms of severity, compared to other disposals for that offence

A list of variables for the MoJ extract of the PNC can be found via this FoI request.

3.2.3 Key Considerations

This is an offence level dataset for individuals. A key benefit of this is that offending histories can be constructed – a key covariate in evaluating individual level outcomes. There are however several weaknesses that evaluators need to be aware of:

- Crime that does not result in identifying an offender will not be recorded in the PNC. This may make the PNC less useful in evaluations that are targeting area level crime reduction, where measures of recorded crime may be more appropriate.
- Personal characteristics recorded in the PNC may be based on officer impressions and may not necessarily be accurate. This may mean that personal characteristics may not match for records that relate to the same individual in the data.
- The capture of individual offender details will depend, to some extent on the targeting of offences and areas by individual police forces, therefore PNC data may be biased as to the types of individuals and areas that are recorded in the dataset.
- There may be some details in the PNC that are missing or are inaccurate; evaluators are encouraged to assess the accuracy and completeness of PNC data before analysing.

3.3 Hospital Episode Statistics

3.3.1 Description

Public health approaches to violence reduction utilises various forms of injury surveillance data which supplement existing criminal justice datasets. Due to variability in the collection of police recorded crime data (under reporting / under recording), health datasets are increasing being used as an additional data source. These include A&E attendance, ambulance call outs and hospital admissions.

The Home Office in consultation with NHS Digital selected Hospital Admissions as a primary outcome measure for monitoring Violence Reduction Units (VRUs).

3.3.2 Available variables

Accident & Emergency (A&E) or Emergency department attendance data (which is accessed through local relationships at a hospital trust level or in aggregate through NHS Digital). Each **Health Episode Statistics (HES)** record contains a wide range of information about an individual patient admitted to an NHS hospital, including:

- patient information, such as **age group, gender and ethnicity**

- administrative information, such as dates and **methods of admission** and discharge
- geographical information such as where patients are treated and the area where they live.

These can be accessed locally (through established relationships with ambulance services or at a hospital trust level), with appropriate information sharing agreements, or at an aggregated level, for example Information Sharing to Tackle Violence (ISTV).

Finished Admissions Episode (FAE)

The reason for admittance is recorded using a cause code using the NHS ICD-10 set of indicators. This supplementary code indicates the external nature of injury. In the case of examination of Violent injury 16 assault codes (ICD-10: X92-Y09) are traditional used. These include assault by bodily force (Y04), Assault by blunt object (Y00) and assault by different types of firearm, (X93-X95) as well as assault by drowning (X92), assault by smoke, fire and flames (X97-X98), and 2 groups of 'Other' (specified (Y08) and unspecified (Y09)) assault mechanisms and a sub-group, hospital admission for violent injury with a sharp object (ICD-10: X99).

Ambulance service call out data

The Ambulance data set (ADS)[45] will contain data items related to:

- Patient demographics (gender, ethnicity, age at activity date);
- Episode information (including arrival and conclusion dates and times, source of referral and attendance category type);
- Clinical information (chief complaint, acuity, diagnosis, investigations and treatments);
- Injury information (data/time of injury, place type, activity and mechanism);
- Referred services and discharge information (onward referral for treatment, treatment complete, streaming, follow-up treatment and safeguarding concerns).

3.3.3 Key considerations

These outcomes are victim focused and are relatively rare, therefore they would not be suitable as evaluation outcomes at the geographies that are typically the focus of interventions. A patient attendance at A&E or hospital may well be outside the local area, and there is variability in the victim being able to clearly identify where an offence took place.

[45] https://www.england.nhs.uk/urgent-emergency-care/improving-ambulance-services/ambulance-data-set/

How an injury was sustained is recorded in the data based on self-reported cause. This may lead to inaccuracies in the data recorded (especially in A&E Attendance where data is collected by a coded based upon the judgement of a receptionist). For example, someone may sustain an injury as part of illegal activity. In cases of domestic violence, victims may be unwilling to report the cause; indeed, abusive partners may accompany victims to hospital to ensure that the nature of the injury is not accurately reported.

Finished Admissions Episode (FAE) data provides greater accuracy due to a clinician's observation of a patient and potential causes of injuries (for example on a ward).

3.4 Recorded crime data (Home Office access route)

3.4.1 Description

The Home Office collect crime data from police forces on reported crimes. This differs from PNC data in that this is not necessarily offender linked – i.e., it includes reports of crimes for which no offender has been identified.

3.4.2 Available variables

The Home Office publish an Annual Data Requirement (ADR)[46] from Police Forces in England & Wales. This document contains; Crime entity Relationships (i.e., how data are linked) and, Crime file specifications focused on location, event, offence detail, person, outcomes etc. For a more detailed list of variables

3.4.3 Key considerations

- See section 3.1.3 for key considerations of local police data.
- The completeness of the data is dependent on data being received from Police Forces; not all police forces may have up to date data at a given point in time.

3.5 Ministry of Justice (MoJ) Data First datasets

3.5.1 Description

The MoJ facilitates the access to linked criminal justice datasets via the Data First initiative,. These datasets include court and probation datasets, but of main interest to YEF evaluators is the access to

[46] Home Office (2023) 2023/24 Annual Data Requirement from Police Forces in England & Wales https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/disclosure-logs/dei-coordination-committee/2023/274-2023-adr-notice-2023-24.pdf

the Police National Computer (PNC) via this route[47]. The PNC can be accessed as a linked dataset with the National Pupil Database (NPD) and other criminal justice datasets (MoJ-DfE Share) – the latter dataset contains educational records linked to the PNC. The details of the PNC records are covered above in section 3.1. - essentially it is a database of offender interactions with the police and CJS, e.g., cautions, arrests, charges, and convictions. It is an offence level dataset linked to an individual.

3.5.2 Available variables

i) Court, prison and probation datasets (including linking datasets): the variables available are detailed on the Data First webpage under 'Datasets'.

ii) PNC and the MoJ/DfE share variables are available on request from datalinkingteam@justice.gov.uk and the list of variables in the National Pupil Database can be found here.

3.5.3 Key considerations

Data access for the PNC via the MoJ is, at first glance, more straightforward than approaching a police force; there is a documented application process a published contact point and assistance available to aid evaluators in accessing the data. However, there are reasons for why local police access to PNC data may be preferred:

- Evaluators may only request PNC in vis the MoJ/DfE datashare if there is an element to the evaluation that relates to education.
- Access to the PNC data via the MoJ is to access the MoJ extract of the PNC. This extract does not contain as detailed information as might be held at the local level. For example, details of the specific locations of offences are more detailed on locally held records than on the MoJ extract.
- There will be a time delay in accessing the MoJ extract. We recommend that evaluators budget from 6 months to a year to access the MoJ extract of the PNC. If evaluators require access to datasets held in the ONS SRS, they should allow additional time to arrange the Assured Organisational Connectivity agreement and to accredit members of the evaluation team (see below) if these are not already in place.
- Key reasons for delay in accessing the data are unsatisfactory information provided on the data application form that results in further queries/requests to amend from the MoJ Data Access Group.

[47] N.B. Data First Initiative also contains other linked datasets: a cross-justice system linking dataset at a person level between all of these six datasets as well as a case linking dataset between criminal courts, prison, and probation datasets, as well as civil and family courts, and via the SAIL Databank, criminal justice datasets linked to Census 2021, with plans for further linkage in future. Offender Assessment data is due to be added imminently.

Accessing the MoJ-DfE Share dataset adds a number of advantages to analysing the PNC alone. Firstly, educational attainment, school attendance and behavior (i.e., exclusion record) are highly predictive of crime outcomes at the individual level. Thus, the use of the linked dataset may significantly increase the statistical power of evaluations if these variables are exploited as covariates. Secondly the NPD may provide a more complete record of individual level offender characteristics that may not be fully captured in the PNC, e.g., detailed ethnicity, language, place of residence over time, etc. There is however a significant draw back in that there is a delay in the linked data being made available for analysis; currently the dataset records offences up to 2020. The MoJ DfE Share dataset is therefore unlikely to be suitable for evaluations that require up to date outcomes.

3.6 National Pupil Database (NPD)

3.6.1 Description

The National Pupil Database (NPD) is a comprehensive set of linked datasets of all individuals educated in English state schools since 2002. While not directly related to offending, aside from education outcomes, The NPD contains individual pupil level outcomes that relate to absence and exclusions; variables that correlate well with concurrent and future offending.

Also included within the NPD is the National Client Caseload Information System (NCCIS) data. This data set is collected by local authorities to report on the activity of individuals aged 16 and 17, for example, to estimate the rate of those Not in Education, Employment and Training (NEET).

3.6.2 Available variables

Details of available variables can be found using the NPD 'find and explore' tool. The absence dataset contains detail on whether the absence is authorised or unauthorised as well as reason for absence. The exclusion dataset also contains details on the timing and the reason for school exclusion. The NCCIS dataset includes current activity at ages 16 and 17; of interest to YEF evaluators, one of the activity codes is 'Custody (Young Adult Offender).

3.6.3 Key considerations

The absence and exclusion outcomes recorded as part of the NPD are usually not directly interpretable as crime outcomes, but they are significant predictors of contemporaneous or future criminal behaviour. Absence data is available termly with a 6–9-month lag; exclusions data is available with a 1-year lag.

The NCCIS data only provides data on whether a young person is in custody at the time of the data collection; any periods of custody outside of this will not be recorded within the dataset. NCCIS data is available in March for the previous academic year.

3.7 Police National Database (PND)

The Police National Database (PND) contains data that relate to investigations – e.g., intelligence. We are not aware of this database being used in evaluation research. The code of practice for access does not preclude access to the PND for non-Police organisations, however, access is strictly controlled and intended for policing usage. A possible use of such a database in evaluation research could be for identifying the effect of interventions on criminal networks.

3.8 OpenSource datasets Police.UK (Single On-line Service) / Office for National Statistics (ONS)

3.8.1 Description

Researchers may utilise police data from other sources for example Police.uk[48] (which is provide by the Police single online service[49]) or the quarterly ONS Crime in England and Wales publication series[50]. These data may be collected and used for either validation or benchmarking purposes. These are particularly useful for geographical analysis; Police.uk incorporates a public facing crime mapping system to promote transparency and accountability[51] and a separate data download and API functionality at a police force and monthly basis at data.police.UK https://data.police.uk/ which contains 13 crime groupings, anti-social behaviour, crime outcomes and stop and search data.

ONS provide two key data releases:

- Quarterly crime statistics covering police recorded crime and CSEW - rolling 12-month figures https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice
- Police recorded crime and outcomes open data tables - quarterly figures by offence types https://www.gov.uk/government/statistics/police-recorded-crime-open-data-tables

These contain multiple datasets which span different levels of offence type and locational aggregation (e.g., Police Forces, Community Safety Partnerships (CSPs)).

[48] https://www.police.uk/
[49] https://www.cds.co.uk/our-work/single-online-home
[50] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice
[51] Chainey, S., & Tompson, L. (2012). Engagement, empowerment, and transparency: publishing crime statistics using online crime mapping. Policing: A Journal of Policy and Practice, 6(3), 228-239.

3.8.2 Available variables

A Police.uk crime dataset contains the following fields;

- Month, Reported by, Falls within, Longitude, Latitude, Location, LSOA code, LSOA name, Crime type, Last outcome category

Police.uk data are aggregated to either areas, a street centroid or at a crime event point level – further information is available on the Changelog and about pages of police.uk.

ONS[52] have produced a *User guide to crime statistics for England and Wales: March 2024* which provides detailed information on the various datasets used to compile crime statistics.

3.8.3 Key considerations

See above for strengths and limitations of recorded crime data.

Police.uk data are highly aggregated across time, crime categories and location to enable anonymity.

- Aggregation - Temporal - Counts by Month.
- Categories - Crime categories e.g., Burglary / Violence & Sexual Offences
- Location - crimes are generally allocated to a street 'centroid' or segment to anonymise data, therefore 'individuals' cannot be identified.

Research by Tompson *et al.* (2015)[53] comparing data accessible through police.uk with corresponding local police provided by the police, identify that these data are sufficient for examination of crime a at lower layer super output area (LSOA), but not at lower geographical levels (streets / postcodes) due to the anonymisation processes.

Not all police forces are submitting information to the single on-line service, due to changes in individual force reporting systems.

3.9 Summary of strengths and limitations of each dataset

The table below presents a summary of some of the strengths and limitations of each of the datasets covered in this guidance document.

These are grouped around the following categories:

[52] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/userguidetocrimestatisticsforenglandandwales

[53] Tompson, L., Johnson, S., Ashby, M., Perkins, C., & Edwards, P. (2015). UK open source crime data: accuracy and possibilities for research. Cartography and geographic information science, 42(2), 97-111.

- Coverage
- Access (lead) time
- Data processing
- Characteristics of data

By way of framing this table, an evaluator needs to take into consideration when various datasets may be used, and the pros and cons of access, processing, and value for evaluation. PNC Data from MOJ is the most comprehensive and definitive dataset for judging (re)offending for evaluations, however, the lead-time for access and approval processes may take a considerable time. Local police datasets are particularly useful, however if an intervention covers multiple police forces, this requires evaluators to have individual negotiations with different (local) data sharing teams, which may require considerable resource in organising. Health data is valuable, but only works for certain evaluations.

Table 3.9: Summary of strengths and limitations of each dataset

Dataset	Route	Strengths	Limitations
Local Police Data			
Police National Computer (PNC)	MOJ access	• National dataset • Most comprehensive dataset for CJS contact / (re)offending • Data extract provided (with pseudo-anonymised data for linking) • Individualised records • Offending history	• Data are subject to some inaccuracies inherent in any large-scale data recording system (e.g., mistyped data entries) • Only covers offenders. • Long lead time for access (12m+) • May be a risk of biased by focusing on different offender / offence types over time and space
	Local Access (e.g., Police, Prisons, Hospitals	• National dataset • Shorter lead time of MOJ • See above.	• Resource required to identify, match and extract data prior to analysis
	Justice Data Lab (JDL)	• JDL creates a matched comparison group based on the characteristics of the intervention cohort.	• Lead time for a report to be provided. • Limited detail – only headline reoffending figures produced for intervention group against comparison group.
Health	Local access	• Hospital admissions / Ambulance Service data will supplement police datasets	• Local datasets at hospital trust level • Some hospital admissions due to crime may not be identified (e.g., domestic violence). •
Local Police Data (recorded crime)	Local access	• Comprehensive - all incidents (calls for service and crimes that are reported to police) • Crime Events and Nominals (Suspects, Offenders, and victims) • Disaggregated / individual events • Geographical detail (x-y coordinates) enables • Ability to link nominals to crime events at locations	• Resource required to identify, match and extract datasets. • Only available at police force level. Therefore, data collection on a force-by-force basis (x43) • Potential for duplicate nominals in police data (if not assigned a PNC number)

MOJ Data First	MoJ – ONS SRS MoJ – SAIL Databank (Not MoJ/DfE Data share)	• Defined access route via MoJ • Other Data First datasets available	• Currently not timely (up to 2021) for PNC outcomes in the MoJ/DfE dataset. •
National Pupil Database (NPD)	DfE – ONS SRS	• Comprehensive; a census dataset that covers everyone educated in state schools.	• Crime outcomes are limited • Absence and exclusions outcomes are not crime outcomes (but are predictive). • NCCIS only records those in custody at the time of data collection.
Single On-line service	Police.uk	• National coverage • Released monthly • Useful for benchmarking activities • Suitable for LSOA level analysis	• Only crime events with limited outcome information • Highly aggregated (crime groups, monthly and to street centroids) and anonymisation process may exclude some crimes. • Not all forces are submitting data to single-on-line service

4 Access procedures

4.1 Local police data

4.1.1 Who to contact

There is not a defined contact for individual police forces; section 2.2 includes advice on how to approach and identify contacts within individual police forces.

4.1.2 Information sharing procedures

Information sharing procedures will differ between Local police forces. A general overview of processes that appear to be common to most police forces is covered in section 2 of this report.

4.1.3 Data access infrastructure

As per section 2, local police data is typically shared with evaluators using Trusted Research Environment model for which ISO accreditation is usually required.

Case Study: Manchester Metropolitan University (MMU) Greater Manchester Police Randomised Control Trial of Hotspot Policing

Manchester Metropolitan University (MMU) had an existing relationship with GMP, developed through previous commissioned projects on data science and violence reduction. MMU had previously established a 2-tier Information Sharing Agreement (ISA) with GMP. There was an overarching agreement and individual schedules for specific datasets and projects. MMU used this ISA and developed a new schedule which was specific for this project. There was already an established secure data infrastructure (IT, servers, research areas and protocols) and GMP were providing data.

Greater Manchester Police (GMP) received Violence Reduction GRIP funding by the Home Office to continue hotspot policing. A condition of this funding was to undertake a randomised control trial (RCT) to enhance the broader evidence base of hotspot policing in the UK. There were limited trials completed in the UK, and most of the evidence was from the United States (Braga et al., 2019).

Two datasets covering a three-year period, were provided for this evaluation: the first was individual crime (event) dataset which included the date and time of the offence, the type of offence, and the location of the crime (i.e., XY coordinates). The second dataset were individual calls for service to the police (Incidents). Again, these data included the date and time, the type and location of the incident. Four variables were constructed from these data. All crime events were used as All Crimes. A sub-set of these were identified as Violent Crimes, utilising the Home Office counting rules and crime-tree. For incidents, all calls for service were used as All Incidents. Those incidents initially coded as Violence were used as Violent Incidents. Using the XY coordinates, the four datasets were mapped using a Geographical Information System (GIS) to identify and examine hotspots across Greater Manchester.

To undertake the randomised control trial, areas with chronic hotspots over three-years were identified. A mapping exercise was utilised to identify hotspot areas using point level data for both incidents (calls for

service) and crimes. The focus of this phase of funding and this RCT would be on residential neighbourhoods, therefore town / city centres were excluded. A 150-metre grid geography was created across Greater Manchester in a GIS. The crime / incident event points were allocated to each of these areas. These 150-metre grids were used to identify chronic hotspot areas in residential settings. For the construction of 'potential' intervention areas, researchers identified three adjacent 150-metre grid cells which had high counts.

This process was repeated across Greater Manchester until approximately 80 locations were identified. These areas were subsequently sense-checked by police officers to understand feasibility of using an area in this intervention. For example, if a school was the epicentre of one of these geographical areas, it was excluded and not used, i.e., the focus on purely residential areas. Eventually through this process, 60 areas were identified. These areas were then randomly allocated into control and intervention areas. They were subsequently clustered into three 'geographical' groups for operational allocation. These were sensible groups, which could be used as a patrol pattern, which managing officers could deploy resources dedicated to the intervention.

The intervention trial took place over a six-month period, where resources were deployed to the 30 intervention areas using a random deployment (shift) pattern, so each site was visited at different times during the day, seven days a week, over course of the intervention. Police officers also completed diaries and took photographs of their deployment and physical presence in intervention locations. Officers also had separate GPS devices to monitor their location across the intervention period, to validate their presence within intervention areas as planned.

Once the intervention period was completed data crime and incident data analysed, so a comparison between the intervention and control sites could be made. Various analytical techniques were used to understand the impact of the intervention on crimes and incidents in hotspot areas.

In addition to this, geographical buffers were made around both the intervention and control areas, and analysis was undertaken to understand both the displacement and diffusion effects of the intervention.

4.2 Police National Computer (PNC) – Ministry of Justice (MOJ) Access

4.2.1 Who to contact

The main point of contact for access to the MoJ extract of the PNC is the data linking team (datalinkingteam@justice.gov.uk)

4.2.2 Information sharing procedures

The data is accessed via application available at gov.uk, though it is expected that evaluators would contact MoJ at the above email address to discuss feasibility and data requirements. Access is granted according to a decision by the MoJ Data Access Governance Board (DAGB) made on the basis of a review of the application by the MoJ Data Access Group (DAG) and their recommendation[54].

Evaluators can use the MoJ access route to obtain data for the analysis of RCTs. In these instances, evaluators would need to supply personal IDs (e.g., first name, last name, DOB, postcode) for matching into the PNC. These would need to be accompanied by a legitimate reason for accessing personal identifying information within the PNC and, in most cases, consent forms from each participant.

4.2.3 Data access infrastructure

The MoJ extract of the PNC is supplied to be accessed at the evaluators own secure setting. The application for this data is the same as for the Data First datasets and should access be granted the data will be provided under a DSA.

Case Study: MOJ Data Request: Greater Manchester Whole System Approach for Women Offenders

Manchester Metropolitan University (MMU) were lead evaluators for the Greater Manchester Combined Authority (GMCA) Whole System Approach for Women Offenders. MMU worked with GMCA to apply to the Ministry of Justice (MOJ) for Police National Computer (PNC) data to explore the proven (re)offending for a cohort of women engaging with the women's centres as part of the intervention. First, women on the intervention were asked to give their consent to access their records from the PNC for a reconviction analysis. A data sharing agreement between MMU, GMCA and the MOJ was established, with the evaluation team providing the following fields from individuals who had consented (Unique local identifier, Individual's Forename, Surname, Date of Birth, Postcode and PNCID (if available)) to enable matching. MOJ then matched these details against the PNC record and provided a data file securely via CJSM. Data included Unique local identifier, Court / caution date, Offence ID, Home Office Offence Code, Disposal Category Date and Duration. These data were stored and analysed in a secure environment before results were approved by MOJ for publication. A range of analyses were completed using the PNC data including 1) proven reoffending using an

[54] In relation to MOJ PNC requests, researchers have needed to select variables from the metadata list for extract. Researchers and evaluators have stated that they have found this process challenging due to a poor understanding to the variables. However, MOJ advise that an initial data ask, is not the final request submission, but part of an ongoing dialogue / process to ensure data sharing principles are met.

offence committed in a one-year follow-up period since first attendance at a women's centre and receiving a court order. This was considerably lower than the re-offending figures for women receiving support from women's centres throughout England. 2) Frequency of Offending - A measure calculated using the 12 months prior to engagement with the women's centres and the 12 month follow up period following engagement with the women's centres. The key lesson for evaluators is to ensure that there is early engagement with MOJ to establish a data sharing agreement due to the lead times for data matching and provision.

Case Study: Sharing data via the local NHS trust: The Solutions Trial

If possible, it makes sense for evaluators to work within existing data sharing agreements as far as possible. This was the case for the YEF funded Solutions trial. This trial is testing an intervention of psychological therapy for those presenting at a custody suite, who are referred to Liaison and Diversion (L&D) teams in the Lancashire and South Cumbria NHS Foundation Trust (LSCFT) region. A set of (secondary) outcomes to be tested in the trial include arrest, caution, reprimands, warnings, and conviction data for participants, outcomes that are collected from the PNC. Instead of accessing the PNC directly, the evaluators are using the fact that there already exists data sharing between i) the PNC and LSCFT, and ii) the local police force and LSCFT. In order to access the necessary PNC data, the evaluators have a DSA between themselves and the LSCFT– i.e., the organisation delivering the intervention. This arrangement avoids the need to negotiate PNC access directly and save resource on developing a bespoke DSA. The key lesson for evaluators is that NHS trusts already have DSA to directly access PNC data and that where interventions involve hospital trusts it makes sense to exploit these rather than develop data sharing agreements directly with the holders of PNC data.

4.3 Justice Data Lab – PNC Reconviction Analysis

4.3.1 Who to contact

For reconviction analysis provided by the Justice Data Lab (JDL) contact justice.datalab@justice.gsi.gov.uk.

4.3.2 Information sharing procedures

The Justice Data Lab is an alternative approach to accessing PNC reconviction analysis. Note that this route is only for *re*-offending outcomes and for individuals aged 14 and over. Using this route to access the reconviction analysis does not involve access to PNC at record level. Instead, evaluators are required to submit personal identifiers of participants to the MoJ who will create a matched comparison group using a defined methodology, and provide a standard reconviction analysis report

and statistics[55]. Confirmation of compliance with GDPR is required as part of the upload of personal identifiers. Full details of using the data lab are here.

4.3.3 Data access infrastructure

There is no specialist infrastructure require for analysis using the Justice Lab – handling of the PNC records are done solely within MoJ. Evaluators will require a CJSM email account to submit personal identifiers and to receive the results – details on how to apply for an account are here.

4.4 Health data

4.4.1 Who to contact

For enquires about health dataset contact individual Hospital or Foundation Trusts.

Certain datasets may also be available by contacting regional organisation who collate data for Trauma and Injury analysis and reporting proposes. For example, the Trauma and Injury Intelligence Group (TIIG) based at Liverpool John Moores University https://tiig.ljmu.ac.uk/ collects data from a range of hospitals across the North West as well as the and the North West Ambulance Service (NWAS).

4.4.2 Information sharing procedures

Access to health data is via application to individual hospital trusts, and are required to meet a lawful basis for processing criminal offence data.

4.4.3 Data access infrastructure

The infrastructure required would be expected to include most of the elements outlined in section 2.8.

4.5 Home Office Recorded crime data

4.5.1 Who to contact

For enquires about access to Home Office recorded crime data

email crimeandpolicestats@homeoffice.gov.uk

4.5.2 Information sharing procedures

There are currently no standardised procedures for accessing Home Office recorded crime data. Data that is currently shared with evaluators is only for those who have been commissioned by the Home Office. These projects are granted access to the HO Recorded crime data by means of a project specific data sharing agreement. These agreements will specify the security requirements for data access

[55] https://www.gov.uk/government/collections/justice-data-lab-pilot-statistics

along the lines of those set out in section 2.7. Access for evaluations that are not commissioned by the Home Office may be possible on a case-by-case basis through contacting the email address above, however, it will be more straightforward in terms of time and likelihood of success to approach local police forces first. The HO is currently reviewing access to recorded crime data for research purposes to better facilitate access for non-HO projects.

4.5.3 Data access infrastructure

As per the above, there is currently no defined route for HO recorded crime access unless part of a HO commissioned project. However, the infrastructure required would be expected to include most of the elements outlined in section 2.8.

4.6 MoJ Data First datasets

4.6.1 Who to contact

In order to access any of the Data First datasets (including the MoJ PNC extract) contact either the Data First team (datafirst@justice.gov.uk) or data linking team (datalinkingteam@justice.gov.uk).

4.6.2 Information sharing procedures

As per access to the MoJ PNC extract (section 4.2.1.2 & 4.2.1.3)

4.6.3 Data access infrastructure

The MoJ Data First datasets are available from the Office for National Statistics Secure Research Service (ONS SRS)[56] and via the Secure Anonymised Information Linkage (SAIL) Databank (though note that the MoJ/DfE data is accessible via this route). In order to access data via the ONS SRS (for MoJ Data First datasets, not including the MoJ PNC extract), evaluators will need to either arrange for their organisation to obtain an Assured Organisational Connectivity agreement with the ONS (details here) or access the data via a SafePod (https://safepodnetwork.ac.uk/). In addition, all individuals who will be accessing the data and/or viewing/discussing unpublished analyses will need to be ONS accredited researchers – details here.

[Please note that the SRS is be replaced by the Integrated Data Service (IDS)].

4.7 National Pupil Database – data

4.7.1 Who to contact

In order to access the National Client Caseload Information System (NCCIS) via the National Pupil Database contact the Department for Education (DfE) at data.sharing@education.gov.uk.

4.7.2 Information sharing procedures

National Client Caseload Information System Data is accessed via application to the Department for Education (DfE) – full details here

4.7.3 Data access infrastructure

Arrangements as per the MoJ Data First datasets (section 4.6.3), i.e., via the ONS SRS.

[56] https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice soon to be replaced by the Integrated Data Service (IDS) https://integrateddataservice.gov.uk/about-the-integrated-data-service

4.8 Single On-line Service / Police.uk

4.8.1 Who to contact

Crime Datasets are available via the https://data.police.uk/ website. The website states that if you have and questions about the data, suggestions for improvements, concerns about the disclosure or personal details, or have noticed any errors in the data, please get in touch with us via the contact form.

4.8.2 Information sharing procedures

Not applicable

4.8.3 Data access infrastructure

No data assess infrastructure requests. Data available under Open Government Licence v3.0

5 Recommendations for evaluators

Practical experience of seeing how the data collected will give evaluators a better idea of the strengths and weaknesses of administrative datasets.

Consider the bias that exists the data – Biases exist in all data, including administrative records such as those recorded in the police national computer (PNC). These biases may lead certain groups to be under or identified as being involved in crime and violence, relative to their true level of involvement. It is important to understand the origins, nature, and extent of these biases at the outset of conducting research using such data and have a plan to address these. Mitigations can include adjustments made to the analysis itself or in how results are reported and contextualized. YEF has a particular focus on racial disproportionality. Children from minority ethnic backgrounds are over-represented in the criminal justice system, particularly Black children, Irish children and children from Gypsy and Irish traveller backgrounds. What leads to this over-representation is a complex mix of individual, society, and system level drivers. Racial disproportionality may also be amplified by biases introduced in the way data is generated and collected. Because of this disproportionality, if we don't challenge the role that racism plays in young people's experiences of youth justice, education and access to employment and mental health support, we won't be able to make the difference we're here to bring about. We encourage evaluation teams to reflect on these issues in their evaluation reports and consider ways to mitigate it: without an acknowledgement of these issues such biases can be perpetuated. A good

starting point to embedding race equity in research is the guide by Child Trends[57]. The report sets out several recommendations including ensuring that evaluation projects have a range of data sources designed to get at the 'root causes' of the phenomenon under investigation, and including children and young people's perspective, based on their lived experiences, when interpreting the data, which may complement the researchers' knowledge and elucidate contextual factors that may influence interpretation of the data.

Relationships are important, and to an extent determine the speed of access. However, high staff turnover is a challenge, as is managing circumstances when evaluation results are unfavourable to the organisation sharing the data.

Build in sufficient lead times for accessing administrative datasets from Police forces and MOJ. The lead time for accessing data from police forces in 3-6 months, and the lead time for accessing PNC data from the MOJ is approx. 12 months.

YEF orientated recommendation - **Funding envelopes / time scales** are too short for delivery of interventions and appropriate follow-up periods for reoffending measures (12months + 6 months = 18 months) is standard evaluation timeline.

Ensure that consent is collected from intervention participants to access Police and PNC data. Appropriate ethical considerations in place to undertake evaluation

Security standards with IT (Cyber Essentials, Vetting. DPIA, ISAs) – need to bring on-board professional services (legal, DP, IT) to facilitate access to these sensitive datasets for evaluation.

Bias and data quality – when establishing and intervention and trial it is important to understand the data quality (strengths and limitations) and if there is any bias with regards to how the intervention will be operationalised.

[57] https://www.childtrends.org/publications/a-guide-to-incorporating-a-racial-and-ethnic-equity-perspective-throughout-the-research-process

6 References

Andrews, K., Parekh, J., & Peckoo, S. (2019) How to Embed a Racial and Ethnic Equity Perspective in Research: Practical Guidance for the Research Process. Child Trends. https://www.childtrends.org/publications/a-guide-to-incorporating-a-racial-and-ethnic-equity-perspective-throughout-the-research-process

Basto-Pereira, M., & Farrington, D. (2019). Lifelong conviction pathways and self-reported offending: Towards a deeper comprehension of criminal career development. *British Journal of Criminology*, 1-18.

Braga, A. A., Turchan, B. S., Papachristos, A. V., & Hureau, D. M. (2019). Hot spots policing and crime reduction: An update of an ongoing systematic review and meta-analysis. *Journal of experimental criminology*, 15, 289-311.

CDS (2024) The Future of Policing https://www.cds.co.uk/our-work/single-online-home

Chainey, S., & Tompson, L. (2012). Engagement, empowerment and transparency: publishing crime statistics using online crime mapping. Policing: A Journal of Policy and Practice, 6(3), 228-239.

College of Policing (2020) Information Sharing https://www.college.police.uk/app/information-management/information-sharing

College of Policing (2023) Police information and records management Code of Practice https://www.college.police.uk/guidance/police-information-and-records-management-code-practice

Curtis-Ham, S., Tompson, L., & Czarnomski, S. (2023). Forewarned is forearmed: the hidden curriculum of working with police crime data. CrimRxiv.

GeoPlace what is a LLPG? https://www.geoplace.co.uk/local-authority-resources/guidance-for-custodians/how-to/about-the-role/what-is-an-llpg

HM Government (2018) Criminal Justice System Exchange Data Standards Catalogue Notification of Change: Introduction of 'Self Defined Ethnicity – 18+1' Standard https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/691544/self-defined-ethnicity-18plus1.pdf

HM Government (2022) Data sharing guidance for researchers seeking permission for secure access to data https://assets.publishing.service.gov.uk/media/62038afa8fa8f510b357cc44/data-sharing-guidance-researchers.pdf

HMICFRS (2018) Crime data integrity programme (crime recording inspections) https://hmicfrs.justiceinspectorates.gov.uk/our-work/article/crime-data-integrity/

HMICFRS (2018) Crime data integrity programme – judgment criteria

https://hmicfrs.justiceinspectorates.gov.uk/our-work/article/crime-data-integrity/crime-data-integrity-programme-judgment-criteria/

Home Office (2021) Crime outcomes in England and Wales: Technical Annex
https://www.gov.uk/government/statistics/crime-outcomes-in-england-and-wales-2020-to-2021/crime-outcomes-in-england-and-wales-technical-annex

HM Treasury (2020) Magenta Book: Central Government guidance on Evaluation
Home Office (2023) 2023/24 Annual Data Requirement from Police Forces in England & Wales
https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/disclosure-logs/dei-coordination-committee/2023/274-2023-adr-notice-2023-24.pdf

Home Office / Ministry of Justice (2023) Data Sharing for the Criminal Justice System Guidance [1]
https://assets.publishing.service.gov.uk/media/652cefa56b6fbf000db7567a/data-sharing-guidance-criminal-justice-system.pdf

Home Office (2024) Home Office Crime Recording Rules for frontline officers & staff
https://www.gov.uk/government/publications/counting-rules-for-recorded-crime

ICO (2023) Data protection impact assessments
https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/accountability-and-governance/data-protection-impact-assessments/#:~:text=You%20must%20do%20a%20DPIA%20before%20you%20begin%20any%20type,or%20serious%20impact%20on%20individuals.

MOJ (2022) The Data First Project: An Introductory User Guide
https://assets.publishing.service.gov.uk/media/62149d4ed3bf7f4f0655016c/data-first-user-guide-version-7.0.pdf

National Cyber Security Centre (2022) Penetration testing: How to get the most from penetration testing
https://www.ncsc.gov.uk/guidance/penetration-testing

National Cyber Security Centre (2024) About Cyber Essentials
https://www.ncsc.gov.uk/cyberessentials

NHS (2022) NHS violence prevention and reduction standard: Guidance notes
https://www.england.nhs.uk/wp-content/uploads/2022/06/B0989-NHS-violence-prevention-and-reduction-standard-guidance-notes.pdf

NHS (2023) ISB1594: Information Sharing to Tackle Violence Minimum Dataset
[1]https ://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1594-information-sharing-to-tackle-violence-minimum-dataset

National Police Chiefs Council (2023) Minimum POLE Data Standards Dictionary
https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/disclosure-logs/dei-coordination-committee/2023/274-2023-pole-data-standards-catalogue-v1.1-1-1.pdf

National Policing Improvement Agency (2011) National Standard for Incident Recording (NSIR) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/116658/count-nsir11.pdf

Office for Statistical Regulation (2024) Administrative data (part 1) https://osr.statisticsauthority.gov.uk/guidance/administrative-data-and-official-statistics/quality-assurance-of-administrative-data-case-examples/administrative-data-part-1/

Office for Statistical Regulation (2023) Systemic Review Outline: Police recorded crime statistics – quality review https ://osr.statisticsauthority.gov.uk/publication/systemic-review-outline-police-recorded-crime-statistics-quality-review/# :~ :text=Police%20recorded%20crime%20statistics%20for,of%20police%20crime%20recording%20practices.

ONS (2021) Census 2021 geographies https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies

ONS (2022) Crime in England and Wales: year ending March 2022 https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmarch2022

ONS (2023) Crime Severity Score (Experimental Statistics) https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeseverityscoreexperimentalstatistics

ONS (2023) User guide to crime statistics for England and Wales: March 2023 https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/userguidetocrimestatisticsforenglandandwales

ONS (2024) Crime and justice https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice

Smith, S., Irving, M., Mann, G., Bjørndal, A., & Lewis, J. (2023) Evaluating Place-Based Approaches: a review of methods used https://youthendowmentfund.org.uk/wp-content/uploads/2023/08/Evaluating-place-based-approaches.pdf

Sutherland, A., Strang, L., Stepanek, M., Giacomantonio, C., Boyle, A., & Strang, H. (2021). Tracking violent crime with ambulance data: how much crime goes uncounted?. *Cambridge Journal of Evidence-Based Policing*, 5(1-2), 20-39.

Tompson, L., Johnson, S., Ashby, M., Perkins, C., & Edwards, P. (2015). UK open source crime data: accuracy and possibilities for research. *Cartography and geographic information science*, 42(2), 97-111.

Thornberry, T.P., & Krohn, M.D. (2000). The self-report method for measuring delinquency and crime. *Measurement and Analysis of Crime and Justice*, 4, 33-83.
University of Cambridge (2020) The Cambridge Crime Harm Index (CCHI) https://www.crim.cam.ac.uk/research/thecambridgecrimeharmindex

Youth Endowment Fund (2021) Data protection information for YEF evaluations Guidance for projects and evaluators https://youthendowmentfund.org.uk/wp-content/uploads/2021/07/YEF-Data-Guidance-Projects-and-Evaluators.pdf

Youth Endowment Fund (2022) Evaluation commissioning guidance https://youthendowmentfund.org.uk/wp-content/uploads/2022/03/17.-YEF-evaluation-guidance-March-2022.pdf

7 Appendices

7.1 Appendix 1: Acronyms

ACPO = Association of Chief Police Offices

BNG = British National Grid

CJS = Criminal Justice Service

CJSM = Criminal Justice Secure eMail

CSEW = Crime Survey of England and Wales

DPIA = Data Protection Impact Assessment

GIS = Geographical Information System

HO = Home Office

JDL = Justice Data Lab

ID = Identification

ISA = Information Sharing Agreement

ISTV = Information Sharing to Tackle Violence

LPD = local police data

LSOA = Lower Layer Super Output area

MOJ = Ministry of Justice

NCCIS = National Client Caseload Information System

NPPV = Non-Police Personnel Vetting

ONS = Office for National Statistics

PNC = Police National Computer

PND = Police National Database

PRIM = Code of Practice on Police Information and Records Management

RCT = Randomised Control Trial

SAIL - Secure Anonymised Information Linkage

SRDM = Self Report Delinquency Measure

So-lS = Single on-line service

SDQ = Strengths and Difficulties Questionnaire

TIC = Taken into consideration

TOC = Theory of Change

URN = Unique Reference Number

VRU = Violence Reduction Units

WEMWBS = Warwick-Edinburgh Mental Well-being scale

YEF = Youth Endowment Fund

7.2 Appendix 2: Stakeholders consulted as part of guidance development

The following stakeholders were consulted as part of the development of this guidance:

- Dr Daniel Acquah (Youth Endowment Fund)
- Dr Nick Axford (Plymouth University)
- Professor Iain Brennan (Hull University)
- Steve Boxford (Cordisbright)
- Professor Simon Coulton (Kent University)
- John Flatley (Home Office) Programme Director Crime & Policing Statistics and Acting Home Office Chief Statistician
- Sukhjit Gill ((Home Office)
- Professor Peter Langdon (Warwick University)
- Mike Parker (South Yorkshire Police / Violence Reduction Unit VRU)
- Kirby Seward (Ministry of Justice MOJ)
- Kevin Wong (Manchester Metropolitan University)

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

This document was last updated in **August 2024.**

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413