EVALUATION REPORT

Youth Endowment Fund (YEF) Administrative Data Guidance

Pilot study report

Mark Ellison and Will Cook

August 2024

PERU Policy Evaluation & Research Unit

Manchester Metropolitan University

YOUTH ENDOWMENT FUND

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 from 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 activities.

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 on 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, Manchester Metropolitan University

Established in 2007, the Policy Evaluation and Research Unit at Manchester Metropolitan University (MMU) 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, 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 following 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 people 'upstream' of involvement in 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 of 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 on 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 provide 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 to relying on the data archive as the only source of access to administrative data on crime and violence. First, the evaluation must be finished and the report published before the data is archived. Second, an approved researcher must have applied to access the YEF data via the Office for National Statistics (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 wants evaluators to access administrative data with crime and violence records within the evaluation period, wherever possible.

This report outlines the administrative data that are available that is likely to be of use to those conducting evaluations of 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 primary or secondary outcomes, 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 (LPD)** includes police-recorded crime data collected by one of the 43 local police forces across England and Wales. LPD includes details of crime events (i.e. offence type, location and date/time) or suspects/offenders demographic information (age, gender, ethnicity)).
- **Police National Computer (PNC) data** is a national dataset which includes information about police cautions and court convictions for individual offenders in England and Wales. The Ministry of Justice (MOJ) receives a data extract to examine offenders' convictions over time and conduct re-offending analyses by offender characteristics.
- **Hospital episode data** includes accident and emergency (A&E) attendance or hospital admission for injuries associated with violence. Data on individual (patient) episodes include demographics (e.g. age, gender and ethnicity).

[1] https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice soon to be replaced by the Integrated Data Service 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 (e.g. MOJ Data First) enable 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 may be collected. Compared to sample surveys, administrative data has a much larger sample size, which 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 the 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 and Thornberry & Krohn, 2000 for discussion). Administrative data is usually restricted in terms of the depth of the variables that are collected and, by its nature, might 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.

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 parents/guardians) consent for their personal records to be

accessed, which may be challenging in some contexts. YEF provides guidance on this in the Data Protection information for YEF evaluations report.[3]

It is important to recognise that such data will naturally present 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 (including a unique reference number [URN], name, date of birth and address) or pseudo-anonymised[4] dataset, 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 (Smith & McVie, 2003), SDQ (Goodman, 1997) or Warwick-Edinburgh Mental Well-being scale (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 for an individual. To enable data matching, it is important that researcher receive a linking variable, such as an identifiable data field (e.g. a URN) or, if not available, a prior pseudo-anonymised reference..

1.2.2 Geographical area/place-based administrative data

In other cases, where interventions seek to reduce the incidence of crime and/or antisocial behaviour, rather than 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 ONS) can either be accessed as administrative or census area–level data; if bespoke geographies or offence types are required, these data can be generated/requested by aggregating up the offence and recorded crime data from precise geographic locations into the desired geography.

Geographical-level data, like individual-level data, does not present a complete picture of crime in an area. This is because large proportions of crime are either not reported to or not recorded by the police. The Crime Survey of England and Wales (CSEW) identified that in 2020, only 42% of comparable

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

[4] 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).

crime incidents were reported to the police (see ONS, 2022[5]). His Majesty's inspectorate of Constabulary and Fire and Rescue (HMICFRS) raised concerns about crime recording by the police (HMICFRS, 2018[6]) which resulted in police-recorded crime no longer being classified as a national statistic designationin 2014 (Office for Statistical Regulation, 2023[7]). This was due to concerns about the quality and consistency of police crime recording practices, with variations 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.[8] Health datasets (e.g. A&E attendance) are now regularly being used to examine violence trends under the Information Sharing to Tackle Violence (ISTV)[9] 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 local police forces, the MOJ, the Home Office (HO), the Single Online 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 to the police (incident) is logged on the local police incident recording system. These incidents may result in a crime is a crime being logged and an arrest made. Information is recorded in 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 into the PNC system as quickly as possible.**[10] If a suspect is charged, they will progress through the criminal justice system, including courts, probation and prison. These organisations collect their own administrative 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 PNC.

Data standards for police operational information and data entities, developed by HO/National Police Chiefs Council, are used to support the consistent and accurate recording of data across the 43

[5] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmarch2022
[6] https://hmicfrs.justiceinspectorates.gov.uk/our-work/article/crime-data-integrity/
[7] 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.
[8] https://hmicfrs.justiceinspectorates.gov.uk/our-work/article/crime-data-integrity/crime-data-integrity-programme-judgment-criteria/
[9] 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
[10] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/873384/PNC_v5.0_EXT_clean.pdf

territorial police forces in England and Wales. The Person, Object, Location, Event standards[11] 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[12]), events (incidents, crimes, episodes) and locations (points, addresses, areas – e.g. census, electoral or administrative aggregated data).

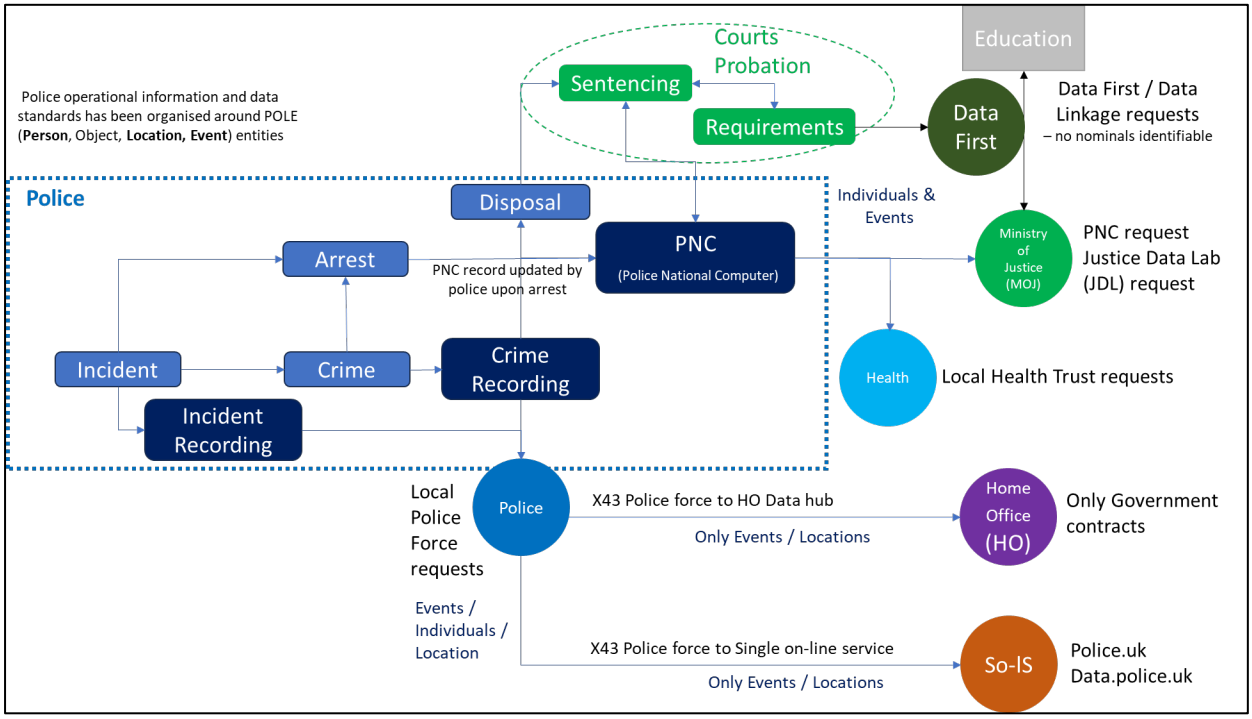

Figure 1: Information flows and access points

1.4 Structure of the 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 practices in operational data management for evaluation.

Section 3 maps the administrative datasets that are likely to be of use for YEF evaluations; these include PNC data, LPD, 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.

[11] 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

[12] "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 victims or witnesses or people not related to offending behaviour, such as missing people or those with licencing violations or road traffic collisions.

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 practices in data access for evaluations.

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.

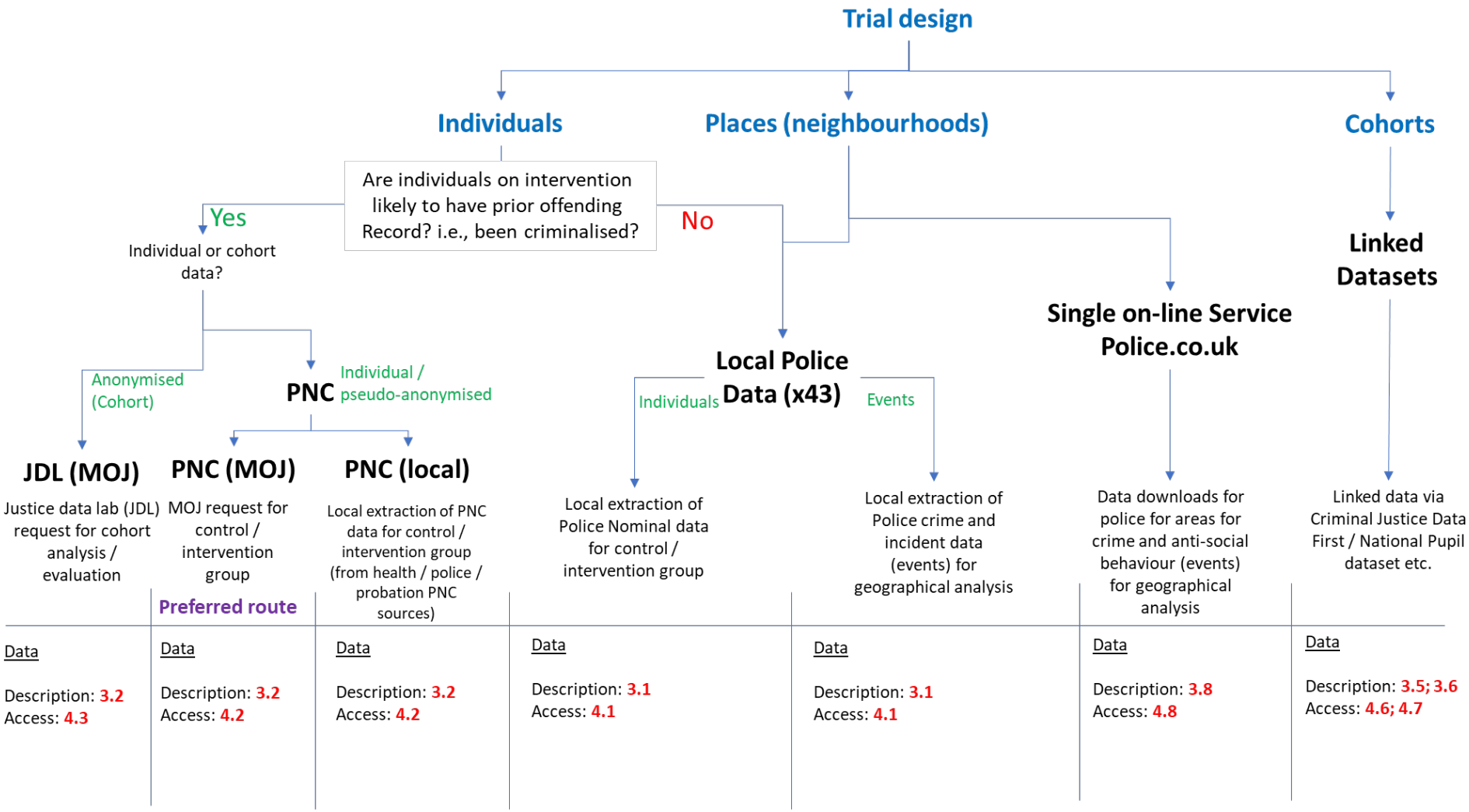

Figure 2: Flow chart of datasets by trial design

Note: Data sections provide signposting to data descriptions (Section 3) and how to access the 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 LPD; however, these steps are also relevant to other datasets. Figure 3A illustrates the stages of pre-data access to the point of ISA sign-off.

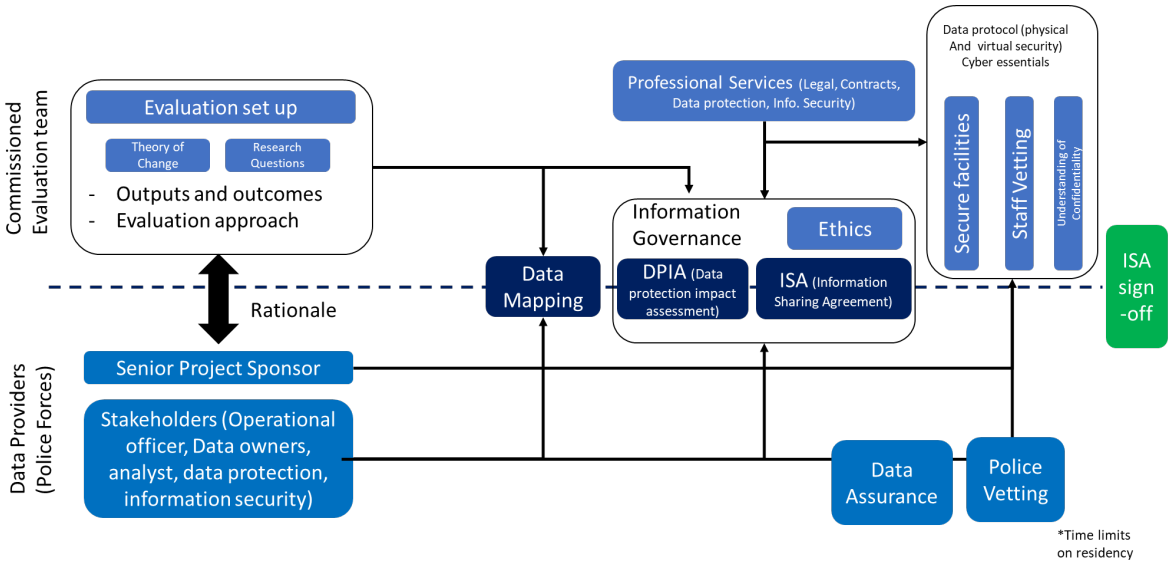

Figure 3A: Pre-data access processes

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

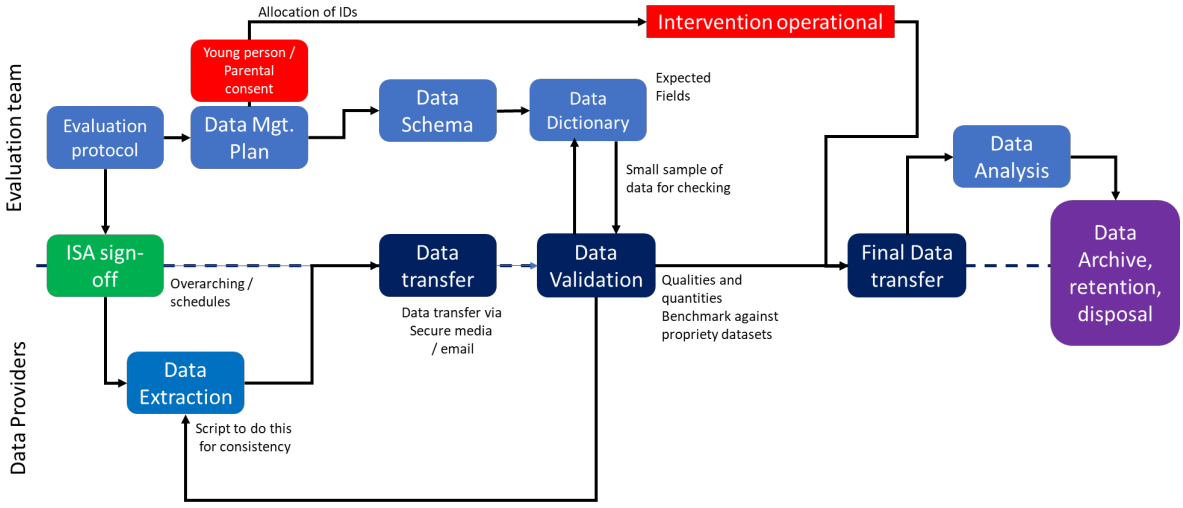

Figure 3B: Operational data management processes

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 practices where there are multiple routes for identifying a point of contact and stakeholders, as well as a section on more direct routes for PNC data from MOJ.

2.2.1 Local police datasets

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

First, it is always good practice to identify a **senior project sponsor,** ideally someone who is at a more senior rank, including senior officers, e.g. Chief Officers, Superintendents or Chief Inspectors. Preferably, this should be someone who works at police headquarters rather than at local police stations. Contacting an external relations, performance, policy or research team is a good start. However, this will vary by police force and prior/existing relationships;[13] establishing these may be challenging and requires persistence. Collaborations, including the N8 policing research partnership (N8PRP, 2024)[14], the Society of Evidence Based Policing[15] and the College of Policing,[16] are a useful starting point for identifying appropriate contacts. Moreover, contacting offices of Police and Crime Commissioners, Community Safety Partnerships and local organisations delivering YEF interventions may also support 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 operational stakeholders 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 or External Relations & Performance) or the team might include a body that acts as a research or academic liaison role.

[13] From consultation with evaluators, many research teams have already established links with individuals in police forces to support the brokering of 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.
[14] https://www.n8prp.org.uk/home/about/
[15] https://www.sebp.police.uk/
[16] https://www.college.police.uk/research/support-research

One of the major issues in working with police stakeholders is their initial identification; in many forces, officers and police staff change roles on a very frequent basis (every 6–12 months). Therefore, over the duration of an evaluation project, researchers may work with different staff in the same role. It is 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 to be briefed appropriately, and a trusting relationship will need to be developed. 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 practices are seen as best practices but with a different set of points of contact and stakeholders. For accessing **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 points of contact. More details are in section 4 of this guidance.

2.3 Data mapping

For any evaluation and project setup, it is vitally important to develop a theory of change (TOC)[17] 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 is operationally independent and has its own crime recording system (however, some forces use the same software supplier). Each police force can advise on the precise data fields and resources required and estimate the time to extract these data. It is important to factor in any resources 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 ISA and data schedules and inform any Data Protection Impact Assessment (DPIA) or ethical requirements.

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

2.4 Developing the Information Sharing Agreement

At this point, the commissioned evaluation team's professional services (for example, legal, contracts, data protection and information security) will be required to advise and support the research team and liaise with the police force data providers' relevant teams, data protection, information security, etc., to draft, formalise and finalise any data-sharing agreements (DSAs) to enable the exchange of data. There are various guidance documents from local police forces, College of Policing,[18] HO,[19] MOJ[20] and NHS[21] on how to do this. YEF have a series of 'comprehensive guidance on the evaluation data archive for evaluators'..

The requirement to provide data may rest upon the approval of secure facilities, secure data transfer processes and 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 start the staff (researcher/support) vetting process early in the project. Evaluation staff may require non-police personnel (NPPV) level 2[22] or level 3[23]. Please be prepared to provide a range of personal information (including details about family, siblings and their relationship, cohabitants, financial information and social media handles). 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 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

Establishing an 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 forces and government websites. For example, The College of Policing website[24] provides a very detailed overview. Individual organisations usually have their own templates, which are then completed jointly, with terms

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

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 various data items.

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 the 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 ISA, individual schedules that are specific to individual projects or datasets, which require particular data processing arrangements, also need to be established.

2.7 Data Privacy Impact Statement

The requirement to conduct a DPIA is set out in the data protection legislation.[25] This may be embedded within an institution's 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 and institutions consider any risks to individuals that are associated with data processing and how to mitigate those risks. It is important that risks should be considered in terms of the likelihood and the severity of any impact on the individuals. More details on DPIAs are available in the YEF data archive. It may be appropriate and more efficient to work with data providers or data controllers to develop a shared DPIA, which can subsequently be utilised by each organisation.

[25] 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, in some cases, a condition of an ISA being agreed upon) 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.[26] Individual organisations may also require external network penetration testing[27] to comply with ISO (International Organization for Standardisation) 27001, the UK Data Protection Act 2018 and the UK General Data Protection Regulation.

For the secure transfer of information, Criminal Justice Secure eMail (CJSM)[28] or Egress Switch[29] can be used to transfer data between people working in criminal justice, public, private and voluntary organisations. The 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 the 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.The preferred method is via Secure USB; however, this may require travel to host organisations (e.g. police HQs or Darlington for PNC). Other approaches evaluators have used include:

1) Digital airlocks, which are fully audited by a Safe Haven[30] team

2) Individual participant files transferred through a whitelist website – drop into airlock into virtual computer two-factor authentication. A trusted Research Environment model for which ISO accreditation is usually required[31].

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

2.9 Operational data management processes

The operational data management process (see Figure 3B) includes important steps for evaluation teams to take that have been identified through best practices (see Section 4).

[26] https://www.ncsc.gov.uk/cyberessentials
[27] https://www.ncsc.gov.uk/guidance/penetration-testing
[28] https://www.cjsm.net/
[29] https://www.egress.com/
[30] https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens
[31] https://ukhealthdata.org/wp-content/uploads/2020/07/200723-Alliance-Board_Paper-E_TRE-Green-Paper.pdf

YEF evaluations require an evaluation protocol.[32] It is also good practice to develop a data management plan for how data are are to be collected, processed and 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 a 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 has been followed and a legal basis for data processing has been established, consent forms, privacy notices and a requirement for data sharing with police forces, health trusts and MOJ for PNC access need to be developed at the start of an intervention. Depending on the individuals within the trial, this may require informed consent and the consent of parents or guardians (see YEF Guidance). It is not possible (very challenging) to retrospectively gain consent from young people and 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 for evaluation, depending on the type and scope of the 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 abilities of 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.

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

- **Data schemas or data dictionaries** – understanding both the quantity and quality of the data – making individual assessments of the completeness of data, where there are limitations and how this may impact sample sizes and statistical power.
- **Data retention or deletion** – aligned to ISA and retention schedules.

More details on the YEF evaluation and data archiving 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

LPD 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 to as LPD in this guidance report. LPD comprises the two main data collections covered in Figure 1.

- LPD on incidents (calls for service) covers all demands placed upon the police requiring assistance, including crime, anti-social behaviour (ASB), 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.
- LPD on crime events and associated nominals (suspects and offenders) is held in crime recording systems. These are used for operational policing purposes. Some of this data are transferred onto the PNC (an overview of the national PNC is included in Section 3.2). Local police records are not a wholly separate data entity from 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 contain more information than the corresponding record on the PNC. However, **LPD 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 LPD will consider one of two dimensions:

- The evaluation will **focus on individuals** receiving an intervention to interrupt or prevent offending (**individual analysis)**
- The impact of 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 dataset. Individuals in a police dataset are generally classed as nominals (and may appear as victims, suspects, offenders or witnesses within a police dataset). 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 that 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 and postcode, 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 ISAs in place) on an individual or cohort level.

The requests 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 include arrests and offending behaviour as either a *suspect* or *offender* and any corresponding disposals which are flagged on the system.

Local police datasets 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 crime events and nominals, which allow data associated with crimes and individuals to be linked. Other data collected includes i) the *modus operandi*, a description of how the crime took place (i.e. the methods and means) and ii) associated flags to provide details about 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 URN and the date of the offence** [committed/reported]

- **The offence type** (which would adopt the HO notifiable offences and counting rules[33])
- **The crime outcome type** (which is linked to the HO Crime Outcome Framework[34])

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 about their disposal. Also, due to the operational nature of these data, individuals may appear as *suspects* in the dataset and then change to *offenders* once a charge/summons is indicated as their disposal.

Type of offence – This code will link to the HO notifiable offences. There are 1,600+[35] 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 crimes include violence against the person, sexual offences, robbery, theft offences, and criminal damage and arson offences.
- Other crimes against society include 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 HO offence codes used in the court proceedings database (these underpin the data used in the Criminal Justice Quarterly statistics publication).

The HO offence codes can also be linked to the Cambridge Crime Harm Index[36] or ONS Crime Severity Score[37] to estimate levels of severity for each crime based on 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 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 for a period prior to and after an intervention.

[33] https://www.gov.uk/government/publications/counting-rules-for-recorded-crime
[34] https://www.gov.uk/government/statistics/crime-outcomes-in-england-and-wales-2020-to-2021/crime-outcomes-in-england-and-wales-technical-annex
[35] https://www.gov.uk/government/publications/counting-rules-for-recorded-crime
[36] https://www.crim.cam.ac.uk/research/thecambridgecrimeharmindex
[37] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeseverityscoreexperimentalstatistics

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 or not in the public interest, where there are any evidential issues, or where the police have stated that their investigation is completed and no suspects have been identified. Only a small proportion of police-recorded crimes (approx. 10%) have suspects identified and assigned as charged/summonsed or TIC or who have been given out-of-court disposals (HO, 2023). In about 40% of crimes, the outcome is assigned as investigation complete – no suspect identified (HO, 2023).

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

Table 1: Key variables in a nominal dataset

Variable	Description	Type	Notes
URN	Unique identifier for each individual		
Date reported	Date crime was reported to police	Date	There are usually multiple dates in LPD: date reported, date recorded, and earliest and latest date committed.
Type of offence	HO offence type	String (code)	This code will link 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	The date of birth is used to calculate the age of the individual.
Gender	Gender of suspect	String	
Ethnicity	Ethnicity of suspect		Standard ethnicity classification (16+1/18+1)[38], which may be self-defined or observed.
Status	Suspect/offender		These are changed retrospectively if a suspect has been identified and has been convicted in court.

[38] 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), which are defined by the intervention, such as area or grids. There are different types of evaluation designs for place-based interventions. There are strong evaluation designs, which include randomisation, and weaker quasi-experimental methods. The types of design listed below[39] apply geographically but can also be used at an individual/cohort level.

Methods involving randomisation include:

- **Cluster-randomised trials** – Areas (or groups) are allocated randomly to intervention or control.
- **Stepped-wedge design** – If all areas will eventually 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 trends and to describe what happens when the trend is interrupted by an intervention.
- **Difference-in-difference** – This 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 person-based approachor not). Comparing their results offers an estimate of impact.
- **Use of concurrent control areas with pre- and post-measurements** – Is a method where and area is enrolled simunlaniously with a treatment area which 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 a propensity score) and those scores can be used in matching.

[39] based on the Magenta Book [HM Treasury, 2020: p13-24], with descriptions by Smith *et al*., 2023 p36–37)

- **Synthetic control methods** – A pool of potential comparable observations using historical data is used to model how areas would have fared without the intervention. The divergence between the actual observations and the synthetic control is 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 the offence and the type of offence (crime code from HO counting rules plus crime tree aggregation). The National Crime Recording Standard sets the expectations for crime recording within the law regarding the timeliness of recording, the need to be 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[40] for addresses (buildings) and other locations (for example, points which represent streets or parks). Data within LPD typically include the x-y coordinates (British National Grid [BNG]), the address, as well as both Census geographies[41] (for example, census wards, LSOAs and output areas) and other ministrative geographies (police beat/wards). These data can be examined using a geographical information system (GIS) to understand crime patterns and can be visualised as thematic or heat maps identifying hotspots.

Table 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 was reported to police	Date	There are usually multiple dates in LPD: date reported, date recorded, and earliest and latest date committed.
Offence	HO offence type	String (code)	This code will link 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	BNG
y-coordinate	Northings coordinate – location of crime	Numeric	BNG
Address	Address of the location of the crime (including postcode for buildings) or location details (e.g. street or park)	String	Address and location qualities may vary between police forces.

[40] https://www.geoplace.co.uk/local-authority-resources/guidance-for-custodians/how-to/about-the-role/what-is-an-llpg
[41] https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies

Special interest markers	A marker to indicate a type of crime (e.g. domestic abuse or knife crime)	String	Markers are used to flag types of crime which cannot be identified through the HO offence types. These are typically used for crimes such as knife crime and domestic abuse, which span multiple HO offence types. Individual police force systems may have on coding systems

Incidents or calls for service

Calls for service on the police (also known as incidents) are reported by the public (999/111) or other emergency services or are observed by the police. The police assess each call for service and make 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] This data includes calls regarding crime-related incidents and non-crime-related incidents, including ASB, public safety and welfare (e.g. domestic incidents, mental ill health, vulnerable persons and missing persons), and road traffic incidents, together with police administration activities and qualifiers (e.g. calls made with good intent by the public, but no was perpetrator present when the police arrived). 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 the details of the initial call for service, how the incident was responded to and the ongoing operational activities by the police and partner agencies.

Table 3: Key variables in an Incident Dataset

Variable	Description	Type	Notes
Incident reference number	Unique identifier for each incident		
Date reported	Date incident was reported to police	Date	Date and time when the incident was logged
Incident type	Broad grouping of incident type	String (code)	These are usually individual crime types: ASB, public safety and welfare, transport, internal administration or qualifiers.
Opening codes	Codes which determine different types of incidents (at the point of the incident being logged)	String (code)	These codes are aligned to the groups identified in the NSIR (see link above), such as domestic disputes or mental ill health.

[42] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/116658/count-nsir11.pdf

x-coordinate	Eastings coordinate – location of the incident	Numeric	BNG
y-coordinate	Northings coordinate – location of the incident	Numeric	BNG
Address	Address of the location of the incident	String	
Closing codes	Codes which determine different types of incidents (once the incident has been resolved)	String (code)	These codes are aligned to the groups identified in the NSIR (see link above), such as domestic disputes or mental ill health.

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.

- HO-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 or 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 or disorder.
- 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 crimes 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 LPD. 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 with an additional unique identifier. However, this is usually undertaken periodically by the police under the Code of Practice on Police Information and Records Management (Police Information and Records Management, 2023[43]), which replaced the Management of Police Information (Management of Police Information, 2005).

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

3.2 Police National Computer data

3.2.1 Description

The 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 and other policing and safeguarding responsibilities. Therefore, the data is reviewed and 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 and vehicles). The MOJ receives 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 those 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 likely of use to evaluators is set out in the tables below. Table 4 includes a list of variables required to request an extract from MOJ PNC. Table 5 includes a list of variables commonly extracted from PNC for analysis and evaluation.

Table 4: Data variables required to undertake a match for PNC extract

Variable	Description	Justification
URN	Unique local identifier	These are necessary to link data back to the original dataset to include local input and outcomes data which are specific to the participant.
Forename	Individual's forename	This is necessary to link the data (due to the potential absence of the PNC ID). Personal details are common between datasets and

[44] https://assets.publishing.service.gov.uk/media/62149d4ed3bf7f4f0655016c/data-first-user-guide-version-7.0.pdf

		are necessary to allow data matching to PNC.
Surname	Individual's surname	As above
Date of birth	Individual's date of birth	As above
Gender	Individual's gender (male, female or unknown)	As above
Postcode	Individual's postcode	As above
PNC ID (if available)	PNC identifier	This is necessary to link to PNC data (if available).

Table 5: Data variables typically provided in a PNC extract for evaluation purposes

Variable	Description	Justification
URN	Unique local identifier	These are necessary to link data back to the original dataset to include local input and outcomes data which are specific to the participant.
Case type	Court or out-of-court disposal	This is necessary to identify the range of disposals given, particularly those resulting in custodial periods.
Court or caution date	Court or caution date	This allows offences to be identified relative to the date of referral to the intervention (i.e. intention to treat).
Offence ID	Number of offences in incident	This allows multiple offences to be grouped into distinct incidents.
HO offence code		This identifies the type of offending.
HO offence category	HO offence category	This identifies the type of offending.
Disposal category	Disposal category for each recorded disposal	This allows disposal history to be identified.
Disposal date	Disposal date for each recorded disposal	This allows disposal history to be identified.
Disposal duration	Disposal duration for each recorded disposal	This allows disposal history to be identified.
Disposal amount	Amount for the first disposal (for fines) and for each recorded disposal where relevant.	This allows disposal history to be identified.
Adjudication code	Guilty/not guilty	Notes: only guilty verdicts 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		This enables the re-categorisation of offence codes.
Disposal rank	Ranking of the disposal in terms of severity compared to other disposals for that offence.	This helps to distinguish between disposals in terms of severity when compared with other disposals for the offence.

A list of variables for the MOJ extract of the PNC can be found at FoI request.

3.2.3 Key considerations

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

- Crimes where the offender is not identified 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 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 utilise 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 increasingly being used as an additional data source. These include A&E attendance, ambulance call-outs and hospital admissions.

The HO, in consultation with NHS Digital, selected hospital admissions as a primary outcome measure for monitoring Violence Reduction Units.

3.3.2 Available variables

A&E or emergency department attendance data is accessed through local relationships at a hospital trust level or in aggregate through NHS Digital. Each **Health Episode Statistics** 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 where they live

These can be accessed locally (through established relationships with ambulance services), at a hospital trust level, with appropriate ISAs, or at an aggregated level, for example, ISTV.

Finished admissions episode

The reason for admittance is recorded using a cause code from the NHS ICD-10 set of indicators. This supplementary code indicates the external nature of the injury. In the case of an examination of a violent injury, 16 assault codes (ICD-10: X92-Y09) are traditionally used. These include assault by bodily force (Y04); assault by a blunt object (Y00); assault by different types of firearms (X93-X95); assault by drowning (X92); assault by smoke, fire and flames (X97-X98); two 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 dataset[45] will contain data items related to:

- Patient demographics (gender, ethnicity and age at activity date)
- Episode information (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 in the geographies that are typically the focus of interventions. A patient's 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 the self-reported cause. This may lead to inaccuracies in the data recorded (especially in A&E attendance, where data is collected by a code based upon the judgement of a receptionist). For example, someone may sustain an injury as part of an illegal activity. In cases of domestic violence, victims may be unwilling to report the cause; indeed, abusive partners may accompany victims to the hospital to ensure that the nature of the injury is not accurately reported.

Finished admissions episode data provides greater accuracy due to a clinician's observation of a patient and the potential causes of injuries (for example, on a ward).

3.4 Recorded crime data (Home Office access route)

3.4.1 Description

The HO collects 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 HO publishes an annual data requirement[46] from police forces in England and Wales. This document contains crime entity relationships (i.e. how data points are linked) and crime file specifications that focus on location, event, offence detail, person, outcomes, etc.

3.4.3 Key considerations

See section 3.1.3 for key considerations of LPD.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 Data First datasets

3.5.1 Description

The MOJ facilitates 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 access to the PNC via this route.[47] The PNC can be accessed as a linked dataset with the National Pupil Database (NPD) and

[46] https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/disclosure-logs/dei-coordination-committee/2023/274-2023-adr-notice-2023-24.pdf

[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.

other criminal justice datasets, such as MOJ-Department for Education (DfE) Share, which 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 linked 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 NPD 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 may be preferred over PNC data:

- Evaluators may only request PNC in via the MOJ-DfE hare if there is an element to the evaluation that relates to education.
- The MOJ extract of the PNC. 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 in locally held records than in the MOJ extract.
- There will be a time delay in accessing the MOJ extract. We recommend that evaluators budget from six 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.
- The key reasons for the 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.

Accessing the MOJ-DfE Share dataset adds a number of advantages to analysing the PNC alone. First, educational attainment, school attendance and behaviour (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. Second, 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 and place of residence over time. There is, however, a significant drawback 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

3.6.1 Description

The 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 dataset 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.

3.6.2 Available variables

Details of available variables can be found using the NPD Find and Explore tool. The absence dataset contains details on whether an absence is authorised or unauthorised, as well as the reason for the absence. The exclusion dataset also contains details on the timing and the reason for school exclusions. 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 six-to-nine-month lag; exclusions data is available with a one-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 for the previous academic year is available in March.

3.7 Police National Database

The Police National Database (PND) contains data that relates to investigations – e.g. intelligence. We are not aware that this database is 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 identifying the effect of interventions on criminal networks.

3.8 OpenSource datasets police.uk (Single On-line service) or Office for National Statistics

3.8.1 Description

Researchers may utilise police data from other sources, for example, police.uk[48] (which is provided by the police So-lS[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 data are particularly useful for geographical analysis: police.uk incorporates a public-facing crime mapping system to promote transparency and accountability[51] Additionally, a separate data download and API functionality, enables researchers to extract data at a police force level, at a monthly basis via data.police.UK (These data are limited to 13 crime groupings, ASB, 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 tables contain multiple datasets which span different levels of offence type and locational aggregation (e.g. police forces and Community Safety Partnerships).

3.8.2 Available variables

A police.uk crime dataset contains the following fields:

[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.

- Month, reported by, falls within, longitude, latitude, location, LSOA code, LSOA name, crime type and last outcome category

Police.uk data is aggregated to either an area, a street centroid or a crime event point – further information is available on the Changelog and About pages of police.uk.

ONS[52] has 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 the 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 or violence and sexual offences
- Location: crimes are generally allocated to a street centroid or segment to anonymise data; therefore, individuals cannot be identified.

Tompson et al. (2015) compared data accessible through police.uk with corresponding local police data and identified that these data are sufficient for the examination of crime at an LSOA but not at lower geographical levels (streets or postcodes) due to the anonymisation processes.

Not all police forces are submitting information to the So-lS due to changes in individual force reporting systems.

3.9 Summary of the 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:

- Coverage
- Access (lead) time
- Data processing
- Data characteristics

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

By way of framing this table, an evaluator needs to take into consideration when various datasets may be used, the pros and cons of access and processing, and the 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 the access and approval processes may be considerable. 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 resources to organise. Health data is valuable but only works for certain evaluations.

Table 6: Summary of strengths and limitations of each dataset

Dataset	Route	Strengths	Limitations
LPD (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 or individual events • Geographical detail (x-y coordinates) enables the ability to link nominals to crime events at locations	• Resource required to identify, match and extract datasets • Only available at the 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)
PNC	MOJ access	• National dataset • Most comprehensive dataset for CJS contact or (re)offending • Data extract provided (with pseudo-anonymised data for linking) • Individualised records • Offending history	• Data is 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 include a risk of bias by focusing on different offender/offence types over time and space
	Local access (e.g. police, prisons and hospitals)	• National dataset • Shorter lead time than MOJ data access request •	• Resource required to identify, match and extract data prior to analysis
	Justice Data Lab (JDL)	• A matched comparison group based on the characteristics of the intervention cohort	• Long lead time for a report to be provided • Limited detail: only headline re-offending figures are produced for an intervention group against a comparison group.
Health	Local access	• Hospital admissions or ambulance service data will supplement police datasets	• Local datasets at the hospital trust level • Some hospital admissions due to crime may not be identified (e.g. domestic violence).
MOJ Data First	MOJ – ONS SRS MOJ – Secure Anonymised Information Linkage (SAIL) Databank (not MOJ-DfE 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

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.
So-lS	Police.uk	• National coverage • Monthly release • 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 the anonymisation process may exclude some crimes • Not all forces are submitting data to So-lS.

4 Access procedures

4.1 Local police data

4.1.1 Who to contact

There is no 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, LPD is typically shared with evaluators using the Trusted Research Environment model for which ISO accreditation is usually required.

Case study: Manchester Metropolitan University Greater Manchester Police randomised controlled trial of hotspot policing

MMU had an existing relationship with Greater Manchester Police (GMP), developed through previously commissioned projects on data science and violence reduction. MMU had previously established a two-tier 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 to this project. There was already an established secure data infrastructure (IT, servers, research areas and protocols), and GMP were providing data.

GMP received Violence Reduction funding from the HO to continue hotspot policing. A condition of this funding was undertaking an RCT to enhance the broader evidence base for 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 an 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. x-y coordinates). The second dataset was individual calls for service to the police (incidents). Again, these data included the date and time and the type and location of the incident. Four variables were constructed from these data. All crime events were used as All Crimes. A subset of these were identified as Violent Crimes, utilising the HO counting rules and the 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 x-y coordinates, the four datasets were mapped using a GIS to identify and examine hotspots across Greater Manchester.

To undertake the RCT, 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 with 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 the 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 to 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 the 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 and to validate their presence within intervention areas as planned.

Once the intervention period was complete, crime and incident data were 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 – Ministry of Justice 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 an application available at gov.uk. However, it is expected that evaluators will 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 made on the basis of a review of the application by the MOJ Data Access Group and their recommendation.[53]

Evaluators can use the MOJ access route to obtain data for the analysis of RCTs. In these instances, evaluators need to supply personal IDs (e.g. first name, last name, DOB and postcode) for matching with 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 from the evaluator's 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 data sharing agreement.

Case study: Ministry of Justice data request: Greater Manchester whole-system approach for women offenders

MMU was the lead evaluator for the Greater Manchester Combined Authority (GMCA) whole-system approach for women offenders. MMU worked with GMCA to apply to the MOJ for PNC data to explore the proven (re)offending for a cohort of women engaging with women's centres as part of the intervention. First, women in the intervention were asked to give their consent to access their records from the PNC for a reconviction analysis. A DSA 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 securely provided a data file via CJSM. Data included a unique local identifier, court/caution date, offence ID, HO offence code, disposal category date and duration.

These data were stored and analysed in a secure environment before the results were approved by MOJ for publication. A range of analyses were completed using the PNC data, including 1) proven re-offending using an offence committed in a one-year follow-up period since first attendance at a women's centre and receiving

[53] 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 of 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.

a court order – this was considerably lower than the re-offending figures for women receiving support from women's centres throughout England and 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 DSA due to the lead times for data matching and provision.

Case study: sharing data via the local National Health Service trust: the solutions trial

If possible, it makes sense for evaluators to work within existing DSAs 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 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, which are outcomes 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 saves resources on developing a bespoke DSA. The key lesson for evaluators is that NHS trusts already have DSAs to directly access PNC data and that where interventions involve hospital trusts, it makes sense to exploit these rather than develop DSAs directly with the holders of PNC data.

4.3 Justice Data Lab – Police National Computer reconviction analysis

4.3.1 Who to contact

For reconviction analysis provided by the JDL, contact justice.datalab@justice.gsi.gov.uk.

4.3.2 Information-sharing procedures

The JDL is an alternative approach to accessing PNC reconviction analyses. 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 the 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.[54] Confirmation of compliance with GDPR is required as part of the upload of personal identifiers. Full details of using the data lab are here.

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

4.3.3 Data access infrastructure

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

4.4 Health data

4.4.1 Who to contact

For enquiries about health datasets, contact individual hospitals or foundation trusts.

Certain datasets may also be available by contacting regional organisations that collate data for trauma and injury analysis and reporting proposes. For example, the Trauma and Injury Intelligence Group 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 North West Ambulance Service.

4.4.2 Information-sharing procedures

Access to health data is via application to individual hospital trusts, and evaluators 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 enquiries about access to HO-recorded crime data email crimeandpolicestats@homeoffice.gov.uk.

4.5.2 Information-sharing procedures

There are currently no standardised procedures for accessing HO-recorded crime data. Data that is currently shared with evaluators is only for those who have been commissioned by the HO. These projects are granted access to the HO-recorded crime data by means of project-specific DSAs. These agreements will specify the security requirements for data access along the lines of those set out in Section 2.7. Access for evaluations that are not commissioned by the HO may be possible on a case-by-case basis by 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 an HO-commissioned project. However, the infrastructure required would be expected to include most of the elements outlined in section 2.8.

4.6 Ministry of Justice 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 the data linking team (datalinkingteam@justice.gov.uk).

4.6.2 Information-sharing procedures

Access is obtained as per access to the MOJ PNC extract (Section 4.2.1.2 and 4.2.1.3).

4.6.3 Data access infrastructure

The MOJ Data First datasets are available from the ONS SRS[55] and via the SAIL Databank (though note that the MOJ-DfE share 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 to be replaced by the Integrated Data Service].

4.7 National Pupil Database – data

4.7.1 Who to contact

In order to access the NCCIS via the NPD, contact the DfE at data.sharing@education.gov.uk.

4.7.2 Information sharing procedures

NCCIS data is accessed via application to the DfE – full details here

4.7.3 Data access infrastructure

Arrangements are made as per the MOJ Data First datasets (Section 4.6.3), i.e. via the ONS SRS.

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

4.8 Single Online 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 any questions about the data, suggestions for improvements or concerns about the disclosure of personal details or if you have noticed any errors in the data, please get in touch with them via the contact form.

4.8.2 Information sharing procedures

Not applicable.

4.8.3 Data access infrastructure

No data assess infrastructure requests are needed. The data is available under Open Government Licence v3.0.

5 Recommendations for evaluators

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

Consider the bias that exists in the data – Bias exists in all data, including administrative records, such as those recorded in the PNC. These biases may lead certain groups to be under- or over-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 to have a plan to address them. Mitigations can include adjustments made to the analysis or 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, societal 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 them – without an acknowledgement of these issues, such biases can be perpetuated. A

good starting point for embedding race equity in research is the guide by Child Trends.[56] 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's and young people's perspectives based on their lived experiences when interpreting the data, which may complement the researchers' knowledge and elucidate contextual factors that may influence the 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 is three to six months, and the lead time for accessing PNC data from the MOJ is approximately 12 months.

YEF-orientated recommendation: Evaluation funding envelopes/time scales provided through YEF commissions are too short to effectively evaluate the delivery of interventions, and the appropriate follow-up periods for re-offending measures (12 months + 6 months = 18 months) are standard evaluation timelines.

Ensure that consent is collected from intervention participants to access police and PNC data and that appropriate ethical considerations are in place to undertake the evaluation.

Security standards with IT (cyber essentials, vetting, DPIAs and ISAs): Evaluators need to bring professional services (legal, data protection and IT) on board to facilitate access to these sensitive datasets for evaluation.

Bias and data quality: When establishing an intervention and trial, it is important to understand the data quality (strengths and limitations) and if there is any bias with regard to how the intervention will be operationalised.

[56] 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. (2024). Forewarned is forearmed: the hidden curriculum of working with police crime data. In The Crime Data Handbook (pp. 9-22). Bristol University Press.
GeoPlace (2024) 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

Her Majesty's Inspectorate of Constabulary and Fire & Rescue Services (2018) Crime data integrity programme (crime recording inspections) https://hmicfrs.justiceinspectorates.gov.uk/our-work/article/crime-data-integrity/

His Majesty's Inspectorate of Constabulary and Fire & Rescue Services (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
https://www.gov.uk/government/publications/the-magenta-book

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
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

Information Commissioner's Office (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

Ministry of Justice (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

National Health Service (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

National Health Service (2023) ISB1594: Information sharing to tackle violence minimum dataset
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 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.

Office for National Statistics (2021) Census 2021 geographies https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies

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

Office for National Statistics (2023) Crime severity score (experimental statistics) https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeseverityscoreexperimentalstatistics

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

Office for National Statistics (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

A&E = accident and emergency

ASB = anti-social behaviour

BNG = British National Grid

CJS = Criminal Justice Service

CJSM = Criminal Justice Secure eMail

CSEW = Crime Survey of England and Wales

DfE = Department for Education

DPIA = Data Protection Impact Assessment

DSA = data-sharing agreement

GIS = Geographical Information System

GMCA = Greater Manchester Combined Authority

GMP = Greater Manchester Police

HO = Home Office

JDL = Justice Data Lab

ID = Identification

ISA = Information Sharing Agreement

ISTV = Information Sharing to Tackle Violence

LPD = local police data

LSCFT = Lancashire and South Cumbria NHS Foundation Trust

LSOA = Lower Layer Super Output area

MOJ = Ministry of Justice

NCCIS = National Client Caseload Information System

NPD = National Pupil Database

NPPV = Non-Police Personnel Vetting

NSIR = National Standard for Incident Recording

ONS = Office for National Statistics

PNC = Police National Computer

PND = Police National Database

RCT = Randomised Controlled Trial

SAIL = Secure Anonymised Information Linkage

SDQ = Strengths and Difficulties Questionnaire

So-lS = Single Online Service

SRS = secure research service

TIC = Taken into consideration

TOC = Theory of Change

URN = Unique Reference Number

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)

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413