Risk of Serious Violence of those already known to the Probation Service

A research study

Ani Setchi, Darren Madgwick, Mary Boeker, Simon Butterworth

Ministry of Justice

Ministry of Justice Analytical Series

2025

Ministry of Justice

Analytical Services exists to improve policy making, decision taking and practice by the Ministry of Justice. It does this by providing robust, timely and relevant data and advice drawn from research and analysis undertaken by the department's analysts and by the wider research community.

Disclaimer

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice (nor do they represent Government policy).

First published 2025

Any enquiries regarding this publication should be sent to us at mojanalyticalservices@justice.gsi.gov.uk

This publication is available for download at http://www.justice.gov.uk/publications/research-and-analysis/moj

ISBN 978 1 911691 89 1

Acknowledgements

This project would not be possible without the extensive resource that two local police forces provided. Thank you to the Data Protection, Data Engineering, Data Analysis and Data Science teams and the many Operational staff at West Midlands Police and Greater Manchester Police that were involved. The time and resource they volunteered in both scoping this project and sharing of data shaped this study and its direction.

We acknowledge the substantial contribution made to the content of this study by our Ministry of Justice colleagues Nick Poyntz, Angela Hawley, Richard Hutchinson and Kimran Rana. We also acknowledge the contribution to the methodology of this study by our Ministry of Justice colleagues Philip Howard, Brian Burton, Robyn Foyster, Michael Williams, Imran Ejaz, Andria Sarri and Talay Cheema, and by the external reviewers of both methodology and content. We also acknowledge the 12 Probation Officers who volunteered their time to take part in user research.

We would like to thank Sapna Sanghvi, Sam Todd, Mandy Tanda, Jamie Daniels, and Duncan Stokes for their direction and ambition to improve public services using data and machine learning, without whom this project would not have happened.

The authors

Ani Setchi, Darren Madgwick, Mary Boeker and Simon Butterworth

Contents

List of tables

List of figures

1. Summary

Predicting the risk of serious violence or homicide has garnered increased worldwide attention, given its implications for public safety, criminal justice, and interventions. There are hundreds of research papers and studies that explore variables associated with serious violence. In this paper, the focus is on work relating to structured actuarial tools and non-sexual serious violence.

This study responds to a 2021 commission when the then-Prime Minister asked the Ministry of Justice to investigate whether improved data sharing between police and probation services could help to improve the Probation Service's ability to assess the risk of people on probation going on to commit serious violence using pre-existing risk assessment tools coupled with richer data sources. The research aims were agreed as part of the formal commission to the Ministry of Justice. These were:

- To review offender characteristics available to the Ministry of Justice that are associated with the risk of serious violence reoffending for people already on Probation.
- To explore the power of local police data to improve the estimation of the risk of serious violence reoffending by people already on Probation.
- To test whether more innovative techniques can improve the risk assessments for future serious violence by people already on Probation.

The cohort, obtained from a snapshot of the prison and probation caseload on a particular date, contained people with at least one conviction and at least one completed risk assessment. It consisted of 274,811 individuals. Of these, 9.3% were women, 90.7% men and 75.8% had spent time in custody. Current age for the cohort ranged from 18 to 99 at the project's snapshot; 21% of the cohort were 25 years old or younger. Based on criminal history data, 64% of the cohort had committed at least one violent offence. Data from two local police forces were used; these two local police datasets had data for 8.7% of the cohort.

By incorporating data from two police forces, and probation and prison systems, this research project explored data snapshots from six data sources. The final dataset contained over 1,500 numerical variables. Several models were tested and compared on this dataset to test different research questions. The main findings were as follows.

- The research found several additional variables that could improve current risk assessments for serious violence, e.g. incidents in custody and new ways of summarising criminal history.
- There are some improvements that could be made to current risk predictors by applying novel data science techniques. Three types of models were tested and, after considering their performance and explainability, a recommendation for future risk assessments would be to consider using Explainable Boosting Machines.
- Despite great effort to collect data from two different police forces, the volume of data was not sufficient to test many hypotheses. There was no evidence that the collected data could significantly improve risk prediction of serious violence by the Ministry of Justice – there was no significant additional value. This could be tested again if more data became available in future research.
- However, this project found evidence that the two police forces involved in the project have some rich data that could be used to create local police risk predictors for serious violence for people with at least one guilty conviction. Such work should be encouraged but it should be considered within ethical and legal standards: the data used should be credible and there should be safeguards in place to deal with bias. Collaboration between forces could greatly increase the possibility for developing such tools; increasing the possible sample size would ensure that tools are fairer for different subgroups of people and would reduce uncertainty within the results.

These findings should be caveated by the fact that hypotheses were tested on a specific cohort, the considered variables were obtained from operational systems where human error is possible in data entry, and some data sources adhered to strict organisational retention rules so could have been incomplete. Every effort was made to consider robust and sensible data processing. Nevertheless, the findings highlight many opportunities and could inform the future direction of operational risk predictors at the Ministry of Justice.

With new technological advances, it is important to consider how risk prediction can move towards better person-specific calculations whilst maintaining sufficient explainability and transparency of models. This research study is part of the Ministry of Justice's commitment to continuous improvement of its actuarial tools and Probation Service offer.

2. Background

Public protection is central to the work of His Majesty's Prison and Probation Service (HMPPS), an agency sponsored by the Ministry of Justice. High-quality public protection relies on accurate risk assessment and the skilled management of those individuals who pose a risk of harm to others.

HMPPS uses several actuarial risk assessment instruments to assess the risk of future proven reoffending. These are applied across HMPPS to inform Pre-Sentence Reports in court, determining suitability for rehabilitative interventions and as part of case allocation. While actuarial risk assessment scores are never the sole determinant of decisions about an offender, they have implications for an offender's sentence plan and risk management.

This publication is one of several pieces of research looking into how the Ministry of Justice can improve current risk assessment instruments. The underlying work was commissioned in 2021 when the then-Prime Minister asked the Ministry of Justice to investigate whether improved data sharing between police and probation services could improve the risk assessments of people on probation going on to commit serious violence. It is a research project that has no immediate impact on any operational or policy decisions relating to the considered cohort, nor the current HMPPS caseload, but could inform future continuous improvement of actuarial risk assessment instruments.

2.1 Motivation for the project

Approximately 600 homicides occur in England and Wales every year (Office for National Statistics, 2025). In addition, some incidents are incidents of significant violence that were only stopped from turning into homicide due to fortunate circumstances, such as fast medical response or wounds being millimetres away from vital organs. Applying the definition of homicide-like offences set out in Section 3.2, the authors estimate that there are over 4,600 convictions for homicide or homicide-like violent offences in England and Wales every year.

The latest available statistics on proven reoffending (Ministry of Justice, 2025) show that adult offenders – those who were released from custody, received a non-custodial conviction at court, or received a reprimand or warning between April and June 2023 - had a guilty reconviction rate of 27.3% in a one-year follow-up period. Of them, 20.9% committed at least one proven reoffence with the index “Violence Against the Person” in that time. Although 20.9% of 27.3% is 5.7%, which is a relatively small proportion, it represents a significant number of offenders given the risk associated with the cohort that probation practitioners manage. It should be noted that homicides and homicide-like offences, as defined in this paper, are a small but immensely harmful subset of offences with the index “Violence Against the Person”.

When offenders on probation are charged with Serious Further Offences (SFO), such as murder, manslaughter and rape, the Probation Service conducts a Serious Further Offence review to determine whether the actions taken by the probation service to supervise the offender were reasonable and defensible and to identify any further action to promote good practice or address any areas for improvement. 579 notifications of SFOs were received for 2022/2023 and as of 30 Sept 2024, 287 of those had been convicted (Ministry of Justice, 2024). These convictions included 60 for murder and 98 for rape and other serious sexual offences. In response to findings from SFO reviews, the Ministry of Justice has implemented action plans to enhance the Probation Service's effectiveness. Whilst more can be done to understand better the risk of committing offences using statistical methods, it is important to emphasise that it is not possible to predict such an offence occurring with any high certainty, only that the risk of occurrence is higher or lower for different offenders.

2.2 Relevant existing tools

The Ministry of Justice is the analytical owner for a suite of operational actuarial predictors, all used within HMPPS’s risk assessment system OASys (Moore, 2015). Continuous improvement, research and validation (Craik et al., 2024, Howard et al., 2023, Horan et al., 2020) ensure the integrity and quality of these tools. They are an important component in practitioner decision making as they reduce individual bias by including an element of consistency and fairness between practitioners. They are always used as part of the overall risk assessment process alongside structured professional judgment.

Risk of serious recidivism (RSR) is an actuarial tool used by HMPPS to help understand the risk of serious reoffending. Embedded within RSR is an actuarial predictor of the risk of serious non-sexual violence (SNSV). RSR-SNSV is a logistic regression model which indicates risk of homicide, wounding and other serious non-sexual violence reoffending. It was last reviewed in 2024 (Craik et al., 2024) and was shown still to perform well for both high and low risk individuals. Another risk predictor used in HMPPS is the OASys Violence Predictor (OVP), which also uses logistic regression to predict both non-serious and serious non-sexual violence reoffending (Howard et al., 2012).

Offender Management staff in HMPPS also use two established structured professional judgement tools within OASys. The Risk of Serious Harm (RoSH) assessment sets a risk level for the risk of serious harm that an individual might pose to others – this professional judgement assessment takes into account the RSR actuarial tool. The Spousal Assault Risk Assessment (SARA v2) assesses the risk that an individual may commit spousal or intimate partner violence and is a framework developed in Canada (Kropp et al., 2015).

The work in this paper is most pertinent to RSR-SNSV as it considers actuarial predictors for serious non-sexual violence reoffending. However, some findings could also have potential relevance to OVP, RoSH and SARA.

2.3 Relevant other Government initiatives

Over the past decade, the Home Office has implemented a multifaceted approach to address serious violence in the UK, focusing on both prevention and enforcement. In 2018, it launched the Serious Violence Strategy (Home Office, 2018), which stressed the importance of early intervention, tackling county lines drug trafficking, supporting community partnerships, and strengthening law enforcement. This strategy led to the creation of Violence Reduction Units (VRUs) in areas most affected by violent crime, fostering collaboration among local partners to address root causes and implement preventative measures. In 2022, the Serious Violence Duty (Home Office, 2022) was enacted, legally mandating public sector organisations to collaborate in identifying and addressing the causes of serious violence within their communities. The duty also requires local areas to develop Strategic Needs Assessments (SNAs) to provide an understanding of how violence is affecting local communities and to help local areas in developing a local Serious Violence Duty Strategy.

The National Police Chiefs' Council and the Association of Police and Crime Commissioners developed the Policing Vision 2030 (National Police Chiefs' Council, 2023), which outlines the need for technology to be central to how law enforcement operates, calling on forces to embrace innovation so that policing can adapt to new threats and opportunities posed by the 21st century. The vision highlights a collaborative, proactive, and digitally-enabled approach to policing, aiming to improve public trust, crime reduction, and the overall effectiveness of law enforcement. It sets the foundation for future reforms and ongoing digital transformation.

The UK government regularly publishes detailed reports on serious violence and homicide statistics to inform policy and public understanding. The Office for National Statistics (ONS) releases annual data on homicide in England and Wales, providing insights into trends, victim demographics, and methods used. The Home Office has published research into the long-term trends and patterns in homicide in England and Wales (Home Office, 2020). A report commissioned by the Home Office (Ashby et al., to be published) provides a more detailed breakdown using data from the Home Office Homicide Index and definitions for homicide sub-types. Additionally, the House of Commons Library provides research briefings, such as the "Homicide Statistics" report (House of Commons Library, 2023), which offers comprehensive analyses of homicide rates across different regions and time periods. The Ministry of Justice publishes official reports on serious violence and homicide statistics in England and Wales. These reports are accessible through the Justice Data website.

Lastly, the Better Outcomes through Linked Data (BOLD) programme at the Ministry of Justice has trialled and is evaluating a Case Information Dashboard, which links and brings key information together from probation, prison and local police databases. This linked dataset provides probation practitioners with an integrated, comprehensive view of individual offender data, facilitating timely, effective decision-making.

2.4 Relevant published research

The prediction of risk of serious violence or homicide has garnered increased worldwide attention, given its implications for public safety, criminal justice, interventions and rehabilitation. There are hundreds of research papers and studies that explore variables associated with serious violence. In this paper, the focus is on work relating to structured actuarial tools that have had significant operational impact. This is justified given extensive research from over 30 countries which suggests that structured, data-driven tools outperform clinical judgment alone in predicting violent reoffending (Viljoen et al. (2024), Wertz et al. (2023), Heilbrun et al. (2010)).

There are many well-established tools, most of which are structured checklists containing empirically derived variables but no underlying weights. For example, HRC-20 (Webster et al., 1997) and HKT-30 (Comité Instrumentarium Forensische Psychiatrie, 2000) provide frameworks in predicting violence for cases with psychiatric disorders. SAPROF (de Vries Robbé et al., 2012) is an additive framework focusing on protective characteristics. DASH (Richards, 2009) is a questionnaire that used to support assessment of the risk of a further domestic abuse offence in the UK.

Actuarial risk predictors are a specific form of structured tools, whereby variables are weighted or otherwise incorporated into a model that outputs a probability score or classification. For legal, ethical and performance reasons, actuarial predictors are most often used in combination with professional judgment. There are many examples of such tools, developed within different data, political, media and ethical landscapes.

LSI-R (Andrews & Bonta, 1995) was developed in Canada and was a risk-need instrument predicting general risk, including violent reoffending. It was later updated to LS/CMI (Andrews et al., 2004), which is still used widely and is validated for men, women, adolescents, and those in custody. Both LSI-R and LS/CMI use a weighted scoring system, which places individuals into Low, Medium and High risk categories. RISC (Adviesbureau Van Montfoort & Reclassering Nederland, 2004), VRS (Wong et al., 2006) and HMPPS's risk assessment system, OASys, have many similarities in approach and validation methodologies to LS/CMI, but they apply logistic regression models within their actuarial tools. FOTRES (Urbaniok, 2004), used in Switzerland and Germany, estimates the baseline recidivism risk for specific offences, including violence. It provides outputs in the form of risk categories but does not provide any details of the underlying algorithm or any probability scores to users. COMPAS (Brennan et al., 2009) is a risk predictor for both general and violent recidivism and is widely used in the US. It uses logistic regression, survival analysis and bootstrap classification methods to produce violence risk categories (1-10 and then Low, Medium and High). COMPAS incorporates gender-specific calibrations and has versions specific to young offenders and for offenders after long periods in custody. It is one of very few algorithms which include features on neighbourhood characteristics and parental criminal history.

Actuarial risk predictors specific to and deployed by local police forces face additional challenges because the data they are trained on, such as emergency calls from the public, non-conviction criminal activities and random-chance interactions with police, could contain biases. RFT (Barnes et al., 2012, and Berk et al., 2008) was used between 2009 and 2017 in Philadelphia, US, to predict serious reoffending, including serious violence, for probation cases. It relied on a random forest algorithm and included statistical reasoning for using charges rather than convictions as predictors. It is another algorithm which included postcode and neighbourhood demographics as variables. In England and Wales there are several projects in development or testing stages, which often take many years due to strict legal, ethical and public scrutiny. Of those that were deployed, the HART algorithm (relevant information can be found in Oswald et al. (2018)) was developed in a joint project between University of Cambridge and Durham Constabulary. It predicted serious reoffending, including violence, using a random forest model. It was used between 2016 and 2021.

More recently, West Midlands Police have been testing a statistical harm score, RFSDi, which is only calculated for people with at least one charge. RFSDi is a calculation of the level of harm associated with current offending. Alongside this, West Midlands Police is also testing a harm model to support Integrated Offender Management, which predicts the likelihood of escalation from low-level to more serious offending. Neither predictor is used operationally yet, but they have passed several rounds of scrutiny by an independent ethics committee; minutes and papers are published regularly.

Other relevant actuarial tools explore the risk of serious violence amongst juvenile offenders (Loeber et al., 2005 and Wolff et al., 2017), amongst sex offenders (Static-99 (Hanson et al., 1999) and VRS-SO (Olver et al., 2018)), amongst individuals arrested for domestic violence perpetration (OXDOV (Yu et al., 2023)), amongst forensic psychiatric patients (VRAG (Harris et al., 1993) and VRAG-R (Harris et al., 2015)) and amongst individuals with schizophrenia-spectrum and bipolar disorders (OXMIV (Fazel et al., 2017)). In addition, there are several actuarial tools that predict general recidivism including CAIS (National Council on Crime and Delinquency, 2018), SIR (Nuffield, 1982), and several state-specific tools in the US, such as ORAS, IRAS and TRAS (Latessa et al., 2010).

The large number of actuarial risk predictors mentioned above suggests a significant development in the risk assessment of violence in the last thirty years. Yet, even when accounting for variations in the number of variables and the types of models, there is no universal or agreed list of variables that is suitable for all operational settings. Differences in data collection, cohort, services and operational application lead to variations in models. Numerical models used within the Ministry of Justice are developed within the requirements and opportunities that the operational data pipelines, systems and procedures allow. They are also scrutinised heavily to ensure legality, ethics, transparency and sustainability are considered alongside performance.

This literature review confirms that current methodologies relating to the risk of serious violence are similar to those applied at the Ministry of Justice: logistic regression models are the most predominant type of model historically, and the themes described by the variables are often similar – criminal history, criminal attitudes, weapon use, emotional control, substance abuse, relationships etc. However, the literature review identified several further research questions for this project. These are:

- Can health, education or employment data, within the limitations of what data is currently collected as part of risk assessments by probation practitioners in HMPPS, be used to improve current risk assessment of serious violence for those under probation supervision? This question was motivated by the several actuarial tools specific to mental health settings, and by questions pertaining to health, education and employment in a large proportion of the mentioned models. Currently, some actuarial tools at the Ministry of Justice already use such data, for example whether someone is currently undergoing psychiatric treatment or is currently unemployed. This research question considers whether other data should be included too.

- Can prison behaviour, based on data from the Prison National Offender Management Information System, be used to improve current risk assessment of serious violence for those under probation supervision? This question was motivated by variables on prison experience in LS/CMI. Currently no such data is used in actuarial tools at the Ministry of Justice.

- Can more novel ways to summarise criminal history, whilst still ensuring simplicity and explainability, improve current risk assessment of serious violence for those under probation supervision? This question was motivated by several models in the literature review that considered Cambridge Crime Harm Index, recency, and length of criminal history when summarising criminal history. Currently, criminal history variables within HMPPS actuarial tools are chosen to include only counting rules that can be performed using mental arithmetic only.

3. Project Definition

All major project decisions, including design, definitions and aims, were agreed after consultation with a Working Group with representatives from the Ministry of Justice, Greater Manchester Police, Greater Manchester Probation Service and the Home Office. In addition, several additional discussions took place with operational staff (representatives from both Greater Manchester Probation and Police Services), policy staff (representatives from Serious Violence and Analysis Unit at Home Office and Public Protection Group at HMPPS), analysts (lead data scientists at Ministry of Justice) and 12 senior probation practitioners as part of user research. These discussions ensured that the research continuously considered the operational and ethical implications of any findings.

3.1 Research Aims

The scope of the research was agreed as part of the formal commission to the Ministry of Justice. The three aims were:

- To review offender characteristics available to the Ministry of Justice that increase the risk of committing serious violence for people already on probation and produce recommendations for new potential variables in future models of serious violence reoffending

- To compare a variety of models to test whether more innovative techniques can improve the estimation of the risk of committing serious violence for people already on probation

- To explore the power of local police data to improve the estimation of the risk of serious violence reoffending by people already on probation. This overarching research question can be split into two more specific questions by considering the operational practicalities around data access and impact:

- Could future Ministry of Justice/HMPPS actuarial risk tools be improved by adding local police data?
- Is local police data predictive? In particular, if local police were to build a risk predictor using the data available to them, how much could their performance be improved by adding Ministry of Justice data?

3.2 Serious Violence Definition

The analytical definition for "serious violence" in this project was agreed after thorough investigation of options based on different offence codes and categorisations by the Home Office, the College of Policing, the Ministry of Justice and a local police force. The agreed criteria was:

- Offences that result in deaths or serious harm with a focus on outcome rather than intent.

The exceptions were offences where offenders present significantly different traits and characteristics, and which would introduce worse accuracy in any data model. These are: (a) Death by drink driving, (b) Death by dangerous/careless/unauthorised driving, (c) Causing serious injury or bodily harm by dangerous/furious driving, (d) Endangering railway passengers, (e) Endangering life at sea, (f) Endangering life within aircraft.

The Home Office and the Ministry of Justice offence codes for "serious violence" are listed in Appendix A. In this project, for brevity, "serious violence" is used to refer to non-sexual violence only. It should be noted that HMPPS uses a pair of actuarial predictors for serious sexual offending as part of its suite of risk assessment tools (Emeagi et al., 2024).

3.3 Observation period

The research project was defined as a cross-sectional study, utilising any known information prior to a cut-off date. Cross-sectional studies examine data gathered at a single point in time within a sample population or predefined subgroup, offering a depiction of the population's characteristics. This decision was made because the research aims to examine several new datasets and data collection is easier for cross-sectional studies than longitudinal ones.

The observation data covered a 9-year period from 01/01/2015 to 01/01/2024. The cut-off date, 01/01/2015, and the length of observation after that date, 9 years, were chosen in the design of the research project to maximise the proportion of people in the cohort who go on to commit a serious violence. The implication of this project decision was that the predictive variables before 01/01/2015 relied on several legacy data systems, which were of potentially lower quality than recent systems. However, this was offset by ensuring the prevalence of the target variable was not too low.

3.4 Cohort

The cohort for the project was constructed based on the availability of data within the Ministry of Justice's operational data pipelines whilst ensuring that the cohort is representative of the probation caseload on a particular date. The cohort consists of all offenders who:

- Had at least one recorded and retained guilty conviction in HMPPS's probation caseload dataset before 01/01/2015
- Had at least one completed full risk assessment in HMPPS's risk management system OASys before 01/01/2015
- Were in custody for less than 90% of the follow-up period between 01/01/2015 and 01/01/2024, and
- Did not die or get deported before 01/01/2024.

The cohort, selected using this definition, consisted of 274,811 individuals. The first two conditions defined the cohort in terms of business need; the latter two provided exclusion conditions to minimise the risk of bias in models. The parameter 90% in the third condition was chosen after analysing the reoffending rates for different options and after consultation with a working group. Of the 274,811 individuals in the cohort, 9.3% were women and 90.7% men, 75.8% had spent time in custody at some point before 01/01/2015, and 15.1% were in custody on this date. Current age on 01/01/2015 for the cohort ranged from 18 to 99, with a median of 33; 21% of the cohort were 25 years old or younger. Based on criminal history data, 64% percent of the cohort had committed at least one violent offence prior to 01/01/2015.

The dataset in this project was constructed at offender level, with each row representing a unique individual. Where multiple records and identifiers existed for some individuals, deduplication was performed to consolidate the information into one record per person. This process followed standard analytical practices at the Ministry of Justice.

3.5 Target Variable

The target variable was a binary flag defined as an offender reoffending and committing at least one serious violence offence between 01/01/2015 and 01/01/2024, allowing for a 9-year period of observation. Only guilty convictions were considered when calculating the target variable. For the chosen cohort, the prevalence was 4.2%, with a prevalence of 4.5% for men and 1.9% for women.

3.6 Datasets

This research project explored data snapshots from six data sources. Only data that would have been known on the cut-off date, 01/01/2015, was collected as independent variables. Only serious violent criminal history in the observation period, 01/01/2015-01/01/2024, was used to define the target variable.

Three of the six datasets were from operational systems used by probation staff in England and Wales. Firstly, Delius. This is the probation service’s case management system. It contains probation supervision information and reports for court. Secondly, OASys. This is a tool to assess the risks and needs of offenders in prisons or subject to probation supervision. Data from OASys contains answers to over 200 questions, which probation staff complete at different points throughout an offender's journey. Thirdly, the Prison National Offender Management Information System (Nomis). This is an operational database used in prisons for the management of offenders. It contains the type of custody, sentence length and involvement in breaches of prison discipline. A total of 586 variables were collected from Delius, OASys and Nomis.

The fourth dataset used in this project was from the Police National Computer (PNC). This is a centralised administrative data source of police information. The PNC system and the PNC data are managed by the Home Office, the College of Policing and the National Police Chiefs' Council. The Ministry of Justice receives an extract of the PNC. This extract contains convictions and cautions for recordable offences in England and Wales, and the criminal history of the offenders with such offences. Some non-recordable offences are also included on the PNC, particularly when they accompany recordable offences in the same case. PNC data provides fuller and more detailed criminal history than is available within Delius, but it does not contain some information known to local police such as dropped charges and callouts. A total number of 312 variables were collected from the PNC for the cohort.

The final two data sources were snapshots from two partnering local police forces: Greater Manchester Police and West Midlands Police. The two organisations were approached to participate in this project, and they agreed despite no direct benefits. The data collected from their systems contained information on dropped charges, incidents that did not result in charges, and police defined flags. The obtained data covered 8.4% of the studied cohort. It was not possible to establish how complete the data was for this subgroup of individuals as offenders often have interactions with police in different regions. However, the data had enough information to test for predictiveness and usability in the context of serious violence risk. A total number of 244 variables were collected from the two police forces for this subset of the cohort. Due to ethical considerations and limited resource constraints, data collected at address level, such as domestic callouts, was not included.

Data linking was achieved using probabilistic matching and manual inspection of records with inconsistencies. Existing data matching protocols were used to link both PNC and local police data. All data was aggregated so that each row represented one person. The data was also anonymised before any modelling commenced in the project.

3.7 Governance

This research project's governance is documented in three Data Protection Impact Assessments (one for the Ministry of Justice, one for Greater Manchester Police and one for West Midlands Police), a Business Case for access to PNC data, and two Data Sharing Agreements. In addition, the project consulted on project decisions with a cross-Government working group and the Ethics Advisory Group at the Ministry of Justice. Feedback was also sought via 12 user research sessions with Senior Probation Officers in one-to-one sessions.

3.8 Media coverage

The Guardian published an article on 8 April 2025 with the headline "UK creating prediction tool to identify people most likely to kill", which generated significant public and media interest. The piece focused on the ethical implications of predictive modelling and its potential impact on civil liberties, with particular attention to fairness, bias, and transparency.

Our lines in response at the time were that the published report presents findings from a research project only — it does not represent an operational tool, nor does it recommend immediate changes to existing practice. The research was designed to test whether data science techniques and broader datasets could improve the assessment of serious violence risk. It includes clear caveats about the limitations and ethical considerations of predictive modelling.

4. Methodology

4.1 Data Engineering

Standard processes were followed when developing a serious violence model. Several methods to define variable interactions and complexities were considered in the data discovery phase. Hypotheses relating to serious violence risk and candidate variables for testing these hypotheses were defined with a range of stakeholders.

Data quality was addressed during this step too, with some variables being discarded, some being imputed using additional information, and some being aggregated. Where inconsistencies in data appeared, significant effort was made to preserve the logic in an offender’s journey, especially for criminal history where there were some inconsistencies between Delius and PNC data. Capping and signed logarithmic transformations were applied in cases where distributions of variables were skewed or had extreme tails.

The final dataset contained over 1,500 numerical variables, with one row per offender. Exploratory data analysis and subgroup analysis revealed that many of the selected variables have valuable information in terms of predictive power.

The dataset was split into training, validation and test sets in a 70:15:15 ratio by stratifying for gender, local police data and target variable prevalence. The test set was kept isolated and only used to evaluate the performance of final models. Where cross-validation was applied, the folds were created from the training and validation datasets only.

4.2 Feature Selection

Superfluous features introduce unnecessary complexity to models which can reduce the model explainability as well as increase the likelihood of overfitting during the training phase. Several methods were tested to select the variables with the most information before training models. These included Principal Component Analysis, methods that considered the target variable (mutual information, Kruskal-Wallis test, Wilcoxon test, point biserial correlation), and methods that considered pair-wise correlation (Cramer's V, Pearson correlation, chi-squared test). These attempts did not yield a process that accounted simultaneously for the predictive power, quality and the cross-correlation of over 1,500 variables so a bespoke multi-step process was created. This process considered the prevalence rates for different feature values, standardised against the number of non-nulls, as well as the Pearson correlation and pair-wise cross-correlation of different features. This process was applied separately for men and women, and then variables were unified to produce a list of 142 variables. The selection was validated by fitting models for randomly chosen sets of 142 variables and comparing the performance on the validation dataset.

4.3 Accuracy metric

The problem was defined as a binary classification problem (whether a person committed serious violence in the observation period or not) with continuous and categorical independent variables. A probabilistic approach to the classification problem was chosen to enhance the interpretability and the comparison between models.

There are three broad groups of model performance metrics for classification models: threshold (those that quantify the classification errors by only looking at the final class predictions versus the actual classes), ranking (those that quantify how effective models are at separating classes for a range of thresholds) and probabilistic (those that quantify the uncertainty in a classifier's predictions). In this project, a decision was made to use a ranking evaluation metric using class labels based on probability predictions. The main reason for this was that this was most relevant to current HMPPS practice, where risk is mainly considered in terms of Low, Medium and High categorisation.

There are two commonly used ranking metrics: the AUCROC and AUCPR (the area under the True Positive Rate vs False Positive Rate curve and the area under the Precision vs Recall curve, respectively). The main difference between these two metrics is that AUC-ROC is symmetric for the binary target variable. AUCPR, on the other hand, applies harsher penalties when predicting the rarer event. Given the highly imbalanced target variable for the dataset (4.2% vs 95.8%), it was more appropriate to use AUCPR as the evaluation metric. A no-skill model, which assigns probabilities randomly from a uniform distribution independently of the features, has an AUCPR of 0.042, the prevalence rate.

4.4 Model Selection

A model discovery was conducted in the open-source machine learning platform H2O. This process narrowed down the choice of models to three: Logistic Regression, eXtreme Gradient Boosting and Explainable Boosting Machine. Hyperparameter tuning was performed for each of these independently.

As part of model selection, several options were considered for splitting the models by (a) gender, (b) presence of custody features, and (c) presence of local police features. These options arose after exploratory data analysis.

For gender, three options were tested: independent models – developing separate models for men and women; a combined model – developing one model treating all rows as equal; and a gender-weighted model – developing one model with some weights to account for the 90.7/9.3 gender split in the cohort. The second option resulted in worst performance for women. The third option required more complex models. The first option had better explainability but relied on a large sample group of women. In this paper, results are presented for the first option, with error bars to visualise the uncertainty due to the larger number of men than women in the cohort.

For presence of custody features, most models handled the nulls for people who had never been to custody relatively well. For logistic regression models, interactions were coded as new variables to account for this. There was not enough local police data to fit a model for the entire cohort. However, analysis of the added value of such data was conducted once a model was already fitted.

4.5 Protected characteristics

There were two protected characteristics that were shown to be good predictors when assessing the risk of serious violence among the chosen cohort: gender and age. This was expected because there is a clear and consistent evidence base, and they are well-established features in current HMPPS actuarial risk predictors.

Several other protected characteristics were processed within the project for the sole purpose of fairness analysis. Variables related to address or offence location information were not processed in any data source.

In addition to protected characteristics, several personal characteristics were processed too. They were used to test existing hypotheses about reoffending. Such information is routinely collected by HMPPS and police to understand each person's risk and needs. This included some special categories of personal data: individuals' mental, physical and psychological health. No other health data was processed in this project.

4.6 Fairness Analysis

Ethnicity data was processed in the project for the purpose of fairness analysis only. The chosen metric for this was the False Positive Rate - a measure of how often a model predicts incorrectly whether someone goes on to commit further serious violence. The False Positive Rates for the five high-level ethnic groups (Asian, Black, Mixed, White, Other) were consistently within the error bars once controlled for differences in prevalence and variable distributions (for example differences in age by ethnicity as described in Office for National Statistics, 2023). In addition, several tests were conducted to check for proxy variables of ethnicity, but no large correlations were observed with any of the model features so the risk of imputing ethnicity within the used data was minimal.

Fairness Analysis was also conducted for gender and current age too, which were the only protected characteristics used in the model. Differences in False Positive Rates for these could be explained with the differences in prevalence and variable distributions. No significant interactions were noted in the developed models between current age and the other variables despite current age having high feature importance. Many significant interactions were observed for gender, which further supports the investigation of separate models for men and women.

5. Findings

5.1 Comparison of models

Three types of models were developed: logistic regression, eXtreme Gradient Boosting (XGBoost) and Explainable Boosting Machine (EBM). These were fit by optimising each model's hyperparameters to maximise AUCPR, measured using fixed K-fold cross-validations with K=5 on a combined training and validation dataset. Cross-validation was used to improve performance and as a tool to check for the stability of feature importance ranking during model training. Each fold was also stratified to have the same proportion of men and women and the same proportion of people with local police data. The best set of hyperparameters was applied to the in-sample dataset, and then on an unseen test dataset.

Using only the 40 variables that added most to the predictive validity (one of the options considered in Section 5.2) and the cohort of men, an EBM achieved an AUCPR of 0.139 ± 0.002, compared with 0.137 ± 0.003 for logistic regression and 0.140 ± 0.002 for XGBoost. The error bars were estimated from the cross-validation folds. A paired t-test on the separate cross-validation values with a 5% significance level determined that the EBM AUCPR was significantly higher than the logistic regression, $t(4) = 5.24$, $p=0.003$. This was also confirmed using a Wilcoxon signed-rank test, $Z=15.0$, $p=0.03$. However, there was no significant evidence that XGBoost outperformed EBM. Similar results were obtained in other comparisons when varying gender and the number of features: XGBoost and EBM performed best, followed by logistic regression but the differences were small. All three performed better than the current operational actuarial tool in HMPPS, RSR-SNSV.

There are advantages and disadvantages associated with any of the three types of models, but EBMs offer an agreeable balance between interpretability and accuracy. EBMs are generalised additive models, which allow for more scrutiny over both results and model assumptions than XGBoost models. Global predictions can be split into additive components, and the response curves can be modified before and after model training to incorporate operational knowledge. Tree-based models such as EBMs are also better at handling correlated variables, with the credit split between them; logistic regression models are not able to do this, so additional data pre-processing is often required with them. Lastly, EBM models can detect pairwise interactions and treat them as separate additive features. This is helpful both in terms of performance improvement but also to detect data issues and for model explainability. Interactions can be introduced manually in logistic regression, but EBMs can find them, not just handle them.

After careful consideration, EBMs were chosen as the primary models in this project for both the men and the women cohorts. This decision is reflected in the recommendations in Section 6.3. The accuracy of these two EBM models depended on the thresholds chosen. The following average performance metrics were calculated for thresholds consistent with those in the current operational actuarial tool in HMPPS, RSR-SNSV. The test data was previously unseen by either model. The similarities in the model accuracies on the in-sample and test datasets indicate that the models were trained in a way that avoided overfitting. The difference in precision between the in-sample and the test datasets for women was expected given the small test set size.

	Men	**Women**
Accuracy (in-sample)	0.933	0.951
Accuracy (test data)	0.935	0.947
Precision (in-sample)	0.200	0.089
Precision (test data)	0.180	0.045
Negative Predictive Value (in-sample)	0.962	0.984
Negative Predictive Value (test data)	0.970	0.963
AUCPR (threshold-independent)	0.139	0.072

Table 1. Average performance metrics for the best EBM models with 40 features for the cohorts of men and women. The threshold in both cases was chosen to be consistent with the one in the current operational actuarial tool RSR-SNSV. The AUCPR performance was compared against the target variable prevalence of 4.5% and 1.9% for men and women respectively.

Serious violence reoffending is difficult to predict; it is a rare event that is predicated on individuals reacting in a certain way to different situations they may or may not encounter in their lives. In addition, the models do not reflect the fact that the riskiest individuals are supervised already or supported through interventions more. For these reasons, such models could never reach perfect precision; rather, they identify individuals who are more likely to react violently to the situations they do encounter.

Whenever a model predicted that someone was more likely to commit a serious violence reoffence, that person did in fact go on to commit the offence roughly 1 in 5 times for men and 1 in 15 times for women (see precision performance in Table 1). For both men and women, the EBM models were approximately 4 times more likely to predict a serious violence reoffence correctly compared to a no-skill model (see AUCPR performance in Table 1 compared to target variable prevalence of 4.5% and 1.9% for men and women respectively) – so the relative improvement in ability to identify risk is similar for men and women despite the smaller sample size for women.

5.2 Number of variables

There are operational costs and data quality risks to consider when using more features in models. Figure 1 shows a comparison of the performance metric, AUCPR, for EBM models trained on different numbers of variables. The prevalence is the rate of serious violence reoffending for the cohort, which is also equal to the AUCPR of a no-skill model – one that assigns a probability randomly from a uniform distribution. Note that the prevalence in women is much lower than it is in men so model performance should be considered in that context. The error bars in both plots represent double the standard errors as estimated from the 5 cross-validation folds. The smaller sample size for women results in significantly more uncertainty, as shown by the larger error bars.

Based on these performance curves, a decision was made to compare EBM models with 40 features (Section 5.1) as this gave the best performance for the fewest features. However, in the following section, feature importance will be calculated using 100 features to give a better overall coverage of feature type with similar model performance.

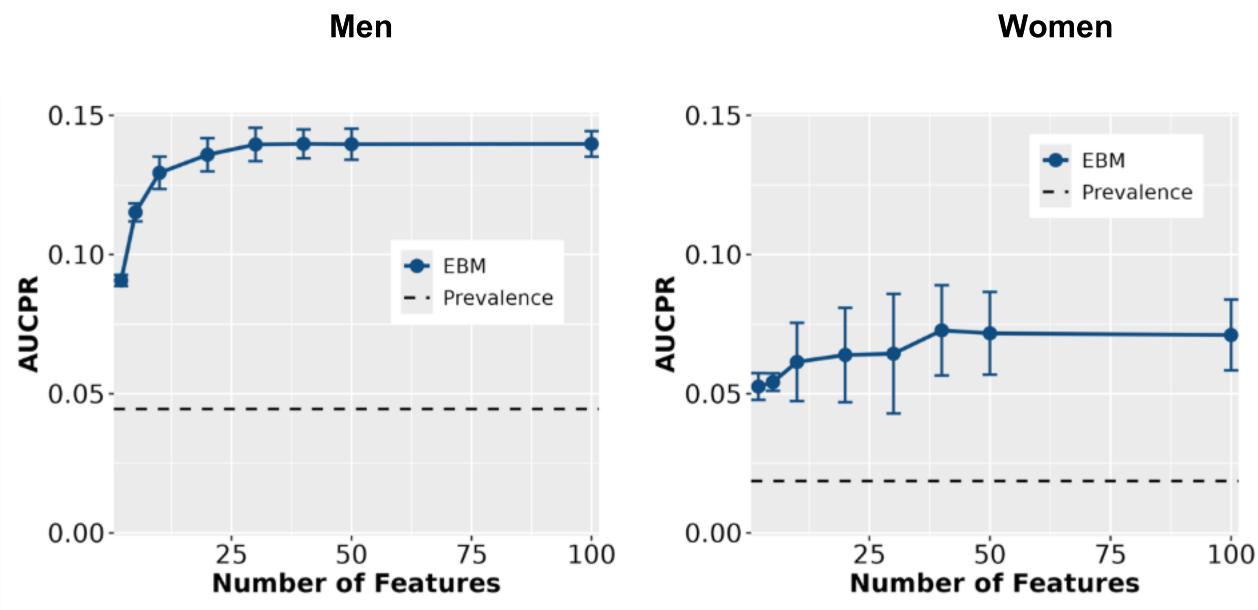

Figure 1. Area under the Precision-Recall curve (AUCPR) for EBM models for men (left) and women (right) for different numbers of features. The prevalence is the proportion of people who go on to commit serious violence. The prevalence is significantly different between men and women. The error bars are calculated as double the standard error on the mean from 5-fold cross-validation.

5.3 Variables that predict risk of serious violence for those already known to the Probation Service

Separate EBM models were trained for men and women, with independent hyper-parameter tuning and automatic feature selection. Variables were aggregated into 14 different groups based on similarity to improve model summarisation. Criminal history was split into three of these groups: one considered only violent offending (Violent terms), one considered counts of any offending (General terms), and one considered specific types of non-violent offending (Other). Custody variables were split into two groups: one group contained variables about behaviour such as incidents, the other all other information about a person’s custodial sentences. The person-specific variables described any personal information that was not captured in any of the other groups. These were psychiatric and psychological problems, emotional issues, issues with attitude to self, learning difficulties, coping difficulties, problem solving skills, abstract thinking, problems with achieving goals, difficulties with work skills, financial management issues, discipline issues, interpersonal skills, impulsivity and temper control. All of these are variables already collected as part of risk assessments in OASys in HMPPS. The variables considered in the group relating to lifestyle and associates held information on peer group influences and number of co-offenders. These variables are also already collected as part of risk assessments in HMPPS; they are specific questions that are linked to risk of reoffending.

Feature importance aggregated into these 14 groups from the best EBM using 100 features for each gender is presented in Figure 2. Feature importance was measured using the average absolute SHAP values. SHAP values indicate the relative contribution of each feature to the model: a higher SHAP value indicates a greater contribution. It should be noted that feature importance does not imply causality.

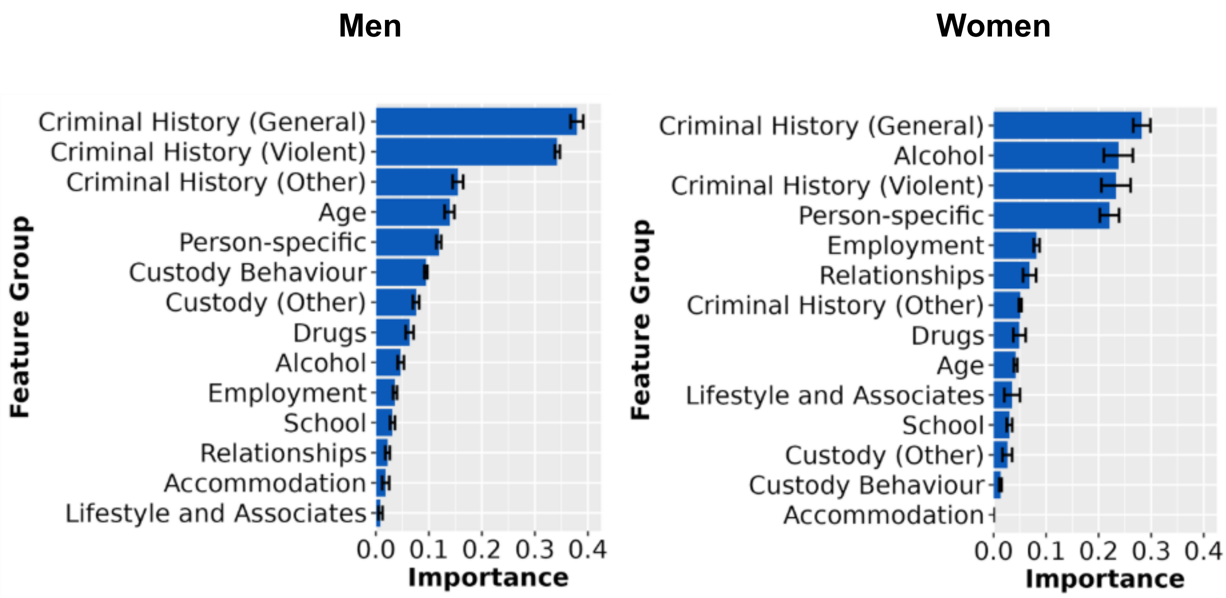

Figure 2. Feature group importance measured by average absolute SHAP (measured in log-odds) for EBM models with 100 features for men (left) and women (right). Individual features have been aggregated into groups before calculating the absolute contributions and averaging across the cohort. Error bars are calculated as double the standard error on the mean from 5-fold cross-validation.

The research suggested several interesting findings. These are summarised by theme with more emphasis on those that are not currently incorporated within HMPPS's current actuarial tool for serious violence, RSR-SNSV. Feature importance within this project identified important factors relating to reoffending rather than primary offending, so should only be considered in that context – other factors were likely to be more important when trying to identify who is more likely to commit a serious violent offence for the first time.

Prison behaviour

An exhaustive search of every feature was carried out to identify the most important predictors of serious violence, using correlations, subgroup analysis (Herrera et al., 2011) and the SHAP feature scores from each of our models. By far the most precise indicators of serious violence for men were recorded incidents of violence in custody – particularly violence against prison staff. For example, men under the age of 25 who assaulted staff in custody were 4 times more likely to subsequently commit serious violence compared with the cohort average. Similar predictive power was also found in violence against non-staff as well as in disorder incidents. There is a question within the risk assessment system OASys on prison behaviour to support probation staff's professional judgement and decisions, but prison behaviour is currently not included in RSR-SNSV. In addition, variables relating to custodial sentences were also shown to be important features, again more so for men than women. These too do not feature in RSR-SNSV, although it is likely that there are proxies within the tool.

A model was fit on the data when Nomis variables were removed. This was done to evaluate the impact that prison behaviour and time in prison have on risk of serious violence reoffending. In this case, using the same methodology, AUCPR dropped to 0.133 ± 0.002. This is a significant reduction compared with the baseline model described in Table 1, which confirms the finding that prison variables are very predictive.

Future work should consider whether data from Nomis could feed into future actuarial tools via automated data pipelines. This recommendation is presented in Section 6.1.

Criminal history variables and the value of PNC as a data source

Previous criminal history (split into general, violent and other) contained very predictive features for both men and women. This concurred with current practices in HMPPS.

Criminal history was extracted from both Delius and PNC for this project. Of these, PNC data contained the more complete criminal history for a person, partly because of (a) containing both convictions and cautions, not just convictions, and (b) longer data retention. When given the choice, EBM models for both men and women chose to use PNC over Delius variables more often. Two of the top three variables for men (the third being current age) and all top three variables for women were PNC variables. These were the rate of violence and general offending over the person's criminal career, age at first sanction, and whether the person had any previous violent offences.

The research in this project supports that PNC data is valuable when assessing the risk of serious violence – an answer to one of the research aims outlined in Section 3.1. However, this project encountered some data quality issues in PNC data, which is expected with any operational and manual data entry system. Possible future work could include using Delius or court data, both available to probation staff, as well as PNC data to capture discrepancies in criminal history across the three systems, with a particular focus on the variables that are shown to be very predictive. In addition, more complex variables describing criminal history (e.g. offences weighted by time, recency, and severity as measured by the Cambridge Crime Harm Index) performed better than simpler counts of offences. This suggests opportunities for other future work in terms of better feature engineering and data pre-processing enabled by expected improvements in digital services for probation staff. This is discussed further in Section 6.2.

Biggest differences between men and women

The biggest differences observed between men and women were in the effect of current age, alcohol-related and relationship-related variables.

Age is a very strong predictor for men – young men have a much higher rate of serious violence than older men – whereas the relationship is significantly weaker for women. These patterns are already incorporated within the current actuarial tool for serious non-sexual violence, RSR-SNSV.

Although alcohol is predictive for both men and women, the effect is much stronger for women in this research project: alcohol variables are the second most significant in the model predictions of serious violence in Figure 2. In addition, issues with relationships are seen to be relatively weak predictors for men, but not for women. Although RSR-SNSV does include several variables relating to both alcohol and relationships, their effect is not as strong for women. Further research on the predictive power of alcohol and issues with relationships for women will be beneficial to test the robustness of this finding.

Health and education data

A research question at the end of Section 2.4 asked whether health, education or employment data could improve risk assessments of serious violence. The person-specific feature group in Figure 2 included some information on health, attitudes and skills – all based only on limited data available to probation practitioners, often self-reported. This group was shown to be important for both men and women.

Variables pertaining to psychiatric and psychological problems were relatively weak predictors. No health-related variables appeared in the top 100 best features for men; psychological problems including depression was the only strong predictor for women – the eighth highest ranked. There were four variables pertaining to psychiatric problems – none appeared in the top 40 best features for women. Without further research opportunities, such as data sharing with health agencies, it is difficult to ascertain whether this is a genuine finding or a result of data with insufficient quality. Similarly, information on education, skills and employment was not found to be very predictive in this research project but it is not known if this is representative of how predictive it could be with better data. Of many considered variables, school attendance issues appeared in the top 40 variables for men, and issues with work skills and school attendance appeared in the top 40 for women. Future work could explore improved methods for collecting such data within existing systems, both by analysing unstructured data better and by encouraging better self-reporting during supervision. The use of health data to predict non-health outcomes could raise legitimate concerns – as such, any changes to data linking or predictive methods would need to be preceded by thorough legal and ethical review.

Dynamic variables

Dynamic variables are defined in this project as those that can change during a person's community sentence or supervision; they exclude new criminal activity and instead focus on risk, needs and circumstance. All dynamic variables in this project, such as alcohol and relationships as discussed above, are answers to questions in the risk assessment tool OASys. For each person in the cohort, the latest answers before 01/01/2015 were considered. Some dynamic variables were shown to be weaker predictors than expected compared with RSR-SNSV, especially for men. For example, issues with accommodation were not chosen by any model in the top 40 by variable importance. A potential reason for this could be that such dynamic variables, as captured in the current risk assessment system, are not updated frequently enough and therefore do not truly capture the subtleties around timing, escalation and change. New risk assessments are usually triggered when a significant event occurs, such as a new appearance in court or a major change in supervision – this means that data on dynamic variables is usually collected weeks, if not months, apart.

One area of further research could be to use natural language processing models to extract structured data from free text data in Delius and Oasys for the purpose of training models and improving the set of risk assessment questions. This was not possible within the current project due to time constraints. This recommendation is mentioned again in Section 6.4.

5.4 Predictive power of local police data

Local police data were sourced from two forces. The acquired datasets covered 8% of the cohort, which made direct comparisons between models difficult. The predictive value of this data was measured in two ways: firstly, by considering how much incorporating it into our existing model could improve predictions for people with local police data; and secondly, by using it to develop a stand-alone model.

The first analysis attempted to improve the predictions of the existing EBM model by adding local police features through subsetting and boosting. No improvement in AUCPR was observed, likely because the local police features were duplicating information already contained in the model. It is possible that the incorporation of other, less correlated information, such as callout data, may result in improvements in future analysis.

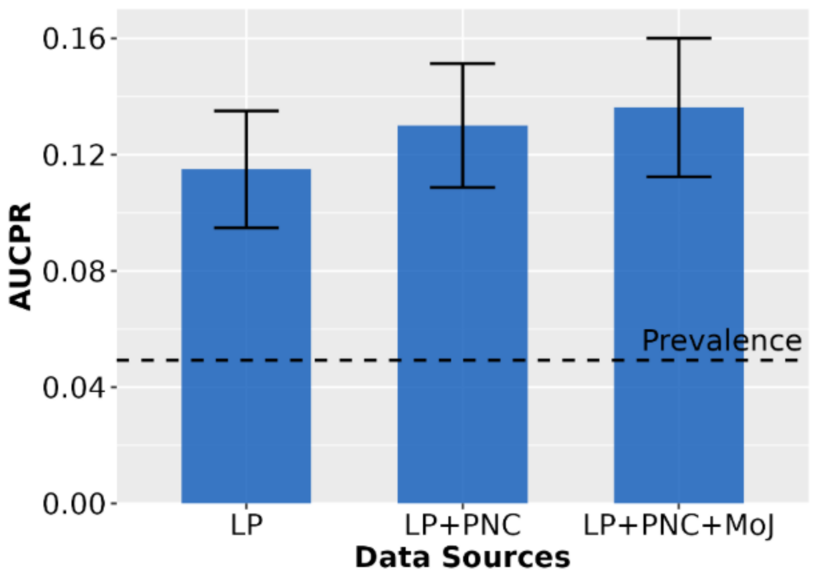

Figure 3. Area under the Precision-Recall Curve (AUCPR) for 40-feature EBM models for men using Local Police (LP) data only (left), a combination of Local Police and PNC data (centre) and all available datasets in this project (right). The prevalence is the proportion of people who go on to commit serious violence and is equal to the AUCPR for a zero-skill model. The error bars are calculated as double the standard error on the mean from 5-fold cross-validation.

Secondly, a new model was constructed, starting from local police data and sequentially adding other data sources (PNC and MoJ respectively). As shown in Figure 3, an EBM model constructed using local police data alone has comparable AUCPR to the EBM trained on all data sources. However, due to the small size of the cohort that had data in either of the two police forces, the error bars are much larger with a subsequent impact on model stability. There was sufficient data, albeit limited, to train a model for men; there was not enough data to do so for women. The variables and the feature importance ranking for the model that used local police variables only are in Appendix B. These are included in this report to evidence that, although 244 variables were tested, which included a wide range of information, criminal history dominated the list. This list is consistent with other findings in this project and is independent of how police forces assess risk currently or plan to do in the future.

The results have four major implications.

- Local police forces possess good-quality data, which can improve how they assess and prevent risk of serious violence. Collaboration between forces can ensure that the sample sizes are large enough to do this effectively for women and produce more stable results for men.

- There needs to be some caution when using these findings because the cohort in this project was chosen to contain only people with at least one guilty conviction. Models developed by local police are likely to have very different cohorts because they do not have bulk automated access to live convictions data. Given that local police, just like the Ministry of Justice, need to adhere to strict ethical standards, a possible way forward would be for local forces to develop and test models using additional PNC data with a condition on the number of guilty convictions being larger than 0. This would evaluate the trade-offs between the benefit and risks of such a model. It could also provide supporting evidence for further research into similar models for people with no convictions.

- The data shared by local police as part of this project was significantly less than expected. This was mainly because of strict retention rules as set out by the Code of Practice for Policing information and Records Management. These rules are good for ensuring information obtained by the police is treated ethically and follow appropriate privacy measures, but they can also hinder research and thus hinder improvements to services. When predicting a rare event, such as serious violence, it is not uncommon to require a long observation period (e.g. 9 years in this project). The features then need to capture behaviours and criminal activity prior to this period, making any feasible project difficult within current retention rules. A possible way forward would be for the College of Policing to allow the storage of anonymised data with the purpose of research only. This would require some additional safeguards and a framework with agreed variables that can cover a varied range of future projects. Nevertheless, the impact to services could be significant.

- Future work would be beneficial if data becomes available from other police forces. Increasing the sample size could confirm the findings in this paper with more certainty, develop a model for women, and test whether new data is predictive. If such further work is conducted, there should be a particular focus on the recency of police data, offenders with criminal history in more than one police regions and the value of callout data.

6. Recommendations

This section lists the main recommendations that have arisen from this research project. Some of these require further research. Others are suggested as long-term ambitions for HMPPS and the Ministry of Justice. These recommendations have been considered carefully and judged against both impact and implementation criteria.

6.1 Recommendation 1: Include prison behaviour within HMPPS risk predictors of serious violence

Custody behaviour is currently included as part of risk assessment but not within current actuarial risk tools. One key recommendation from this research is that recorded incidents in prison should be included in future violent risk predictors and priority should be given to collecting this data, where it is not already. Further work could be conducted to automate the data pipelines from custody datasets to risk assessment digital tools, and more testing could be done to ensure the robustness and integrity of this data.

6.2 Recommendation 2: Provide Probation practitioners and Prison Offender Managers with the tools to summarise criminal history more effectively and with fewer errors

Previous criminal history is important when predicting future risk. Currently, collecting data on previous criminal history is done by probation practitioners requesting a PNC data extract for a person and manually counting the number of offences according to specific rules. This process can be time consuming and subject to human error, but it also limits the possibilities in terms of feature engineering. The Ministry of Justice’s ambition to improve Digital Services for frontline staff provides the opportunity to test new tools that can help in this domain. Improved data pipelines and modernising the general data landscape could improve risk assessment accuracy and save probation staff time. Possible future work could involve a data scraper, summariser, discrepancy finder or a feature engineering tool. Such a project could also improve data quality, introduce new data pre-processing and improve explainability within risk predictions.

6.3 Recommendation 3: Use new tools in the next generation of risk predictors

The current risk predictor for serious violence in HMPPS, RSR-SNSV, is based on a logistic regression algorithm and around 40 variables. This is an excellent operational tool because it provides explainability and has sufficient performance accuracy. Recent developments in machine learning research, however, provide alternative methods, which match this explainability but offer slight improvement in performance. This report has demonstrated how novel techniques, such as Explainable Boosting Machines, can improve the assessment of the risk of serious violent reoffending. In addition, current technologies could soon allow for more frequent and easier-to-implement updates so that the model can change in line with new research. Such continuous improvement and development, together with integrated monitoring and evaluation, should provide a good basis for future risk prediction models within HMPPS services.

6.4 Recommendation 4: Improve the collection of dynamic variables to inform risk assessments

Dynamic variables in the risk predictions of serious violence were extracted based on the latest available OASys risk assessment answers. In many cases, the latest assessment was conducted many years before the 9-year observation period and, therefore, is not a true representation of a dynamic variable. Future work to explore this topic, subject to appropriate ethical considerations, might include:

- New feature engineering that considers recency information – this would allow to quantify the importance of when certain events have happened, not just that they have happened.
- Use of statistical models and natural language processing to collect dynamic data from free text fields – this would widen the field of variables that can be considered within a potential model but can also improve data collection and recency for current variables.
- Monitor models against live data and incorporate notifications and triggers when significant changes in dynamic variables are found by the models.

Such work should be possible in the future thanks to the digital improvements that the Ministry of Justice are expected to deliver to probation practitioners soon as part of the Ministry of Justice Digital Strategy 2025 (Ministry of Justice, 2022).

6.5 Recommendation 5: Further research into new health and education data

There is no formal way to validate health and education data that is collected as part of risk assessments in HMPPS – most is self-reported or relies on probation practitioners' enquiries. Future work could explore improved methods for collecting such data within existing systems, both via analysing unstructured data better and by encouraging better self-reporting during supervision. The use of health data to predict non-health outcomes could raise legitimate concerns – as such, any changes to data linking or predictive methods should be preceded by thorough legal and ethical review.

6.6 Recommendation 6: Further work is required to understand the value of local police data in probation risk assessment processes

One of the main points of interest for this project was to identify whether local police data could significantly improve how probation staff predict risk of serious violence – a process that would require significant transformation of data pipelines and data access pathways.

Based on data for only 8% of the cohort, this project was not able to verify that local police data could improve an actuarial risk assessment for serious violence. However, further work with more data from other police forces could be conducted to verify this conclusion and to test against other types of reoffending risk, not just serious violence. Any finding that does verify the potential to improve predictive validity would warrant detailed ethical and legal consideration about its application to probation risk assessment processes, not all of which apply such data in the deterministic way that actuarial risk assessment instruments do.

Many probation staff, as part of user research, described how the context and the timeliness of police data can make a significant difference to the success of delivery of sentence planning, interventions and case management within probation rather than predicting further offending. Access to weekly local arrests, domestic callouts and weapon or violent incidents can be extremely valuable to this end. Some areas in England and Wales do already have such regular data feeds; this can be encouraged nationally.

6.7 Recommendation 7: Local police have rich data and could develop their own predictive tools with some considerations in place

The Ministry of Justice is grateful to the two police forces that were involved in this project, Greater Manchester and West Midlands. Engagement with colleagues in these forces has identified the many strands of work and ambition to improve the prediction of risk of serious violence there.

The findings in Section 5.4 suggest that police forces hold useful information when assessing the risk of serious violence. These findings are limited to a cohort where everyone had at least one guilty conviction so should only be considered within that context. The variables that were found to be predictive are dominated by ones describing criminal history – this is both intuitive and reassuring. Local police having data, which is rich enough to assess the likelihood of serious violence and support professional judgement, brings a challenging ethical dilemma. In some situations, acting on information based on non-convictions, such as numerous incidents where no further action was taken, could have prevented serious violence, whereas in other situations acting on the same information could be seen as unfair.

Local police forces face many challenges when assessing risk of violence because of (a) smaller sample sizes compared to national cohorts, (b) ethical challenges around data on people with no convictions, and (c) public mistrust of how data is collected, given the nature of it is often subjective (police officers' judgements of situations). Models trained on such data can reinforce and propagate bias, especially in cases where people in the data have no guilty convictions. One possible route forward is for local police forces to link their data to PNC data, to which they should be able to gain access, to create models for people with convictions only.

The Ministry of Justice, the Home Office and the College of Policing could be more active in facilitating more collaboration between forces to ensure that models are developed on larger sample sizes. This would minimise the risk of data variance, if that is a concern, would improve public trust in models and would allow for models that are fair for subgroups of the population.

6.8 Recommendation 8: Local police should have access to legacy datasets for research purposes

This project shows that there is good, predictive data within local police datasets to predict serious violence. However, strict retention rules, as set out by the Code of Practice for Policing information and Records Management, can hamper research into rare events where volume and longitude of data is essential. Exceptions for research purposes can be made here that still comply with the operational implementation of the rules, for example, storing anonymised datasets before archiving data or allowing retention of data on servers that cannot be accessed by operational staff. This requires strategic ambition and resourcing but could facilitate great insight and potential improvement to public services.

7. Conclusion

This project considered the following research aims.

The first was to review offender characteristics available to the Ministry of Justice, including PNC data, that increase the risk of committing serious violence for people already on probation and to produce recommendations for new potential variables in future models of serious violence reoffending. The research found several such additional variables including incidents in custody and new ways of summarising criminal history.

The second was to compare a variety of models to test whether more innovative techniques can improve the estimation of the risk of committing serious violence for people already on probation. Three types of models were tested and, after considering their performance and explainability, a recommendation for future risk assessments would be to consider using Explainable Boosting Machines.

The third was to explore whether future Ministry of Justice/HMPPS actuarial risk tools could be improved by adding local police data. Despite great effort to collect data from two different police forces, the volume of data was not sufficient to test many hypotheses. There was no evidence that the data collected could significantly improve risk prediction of serious violence by the Ministry of Justice. This could be tested again if more data became available in future research.

The final aim was to assess whether local police data is predictive in the context of serious violence proven reoffending. This project found evidence that local police forces have some rich data that could be used to create local police risk predictors for serious violence reoffending for people with at least one guilty conviction. Such work could have great impact and benefit, but it should be considered within ethical and legal standards; the data used should be credible and there should be safeguards in place to deal with bias.

All findings should be caveated by the fact that hypotheses were tested on a specific cohort, the considered variables were obtained from operational systems where human error is possible in data entry, and some data sources adhered to strict organisational retention rules so could have been incomplete. All effort was made to consider robust and sensible data processing. Nevertheless, the findings highlight many opportunities and could inform the future direction of operational risk predictors at the Ministry of Justice.

With new technological advances, it is important to consider how risk prediction can move towards better person-specific calculations whilst maintaining sufficient explainability and transparency of models. This research study is part of the Ministry of Justice’s commitment to continuous improvement of its actuarial tools and its Probation Service offer.

References

Adviesbureau van Montfoort, & Reclassering Nederland (2004). RISc versie 1.0. Recidive Inschattings Schalen. Handleiding. [RISc version 1.0. Risk Assessment Scales. Manual] Utrecht: Reclassering Nederland.

Andrews, D. A., & Bonta, J. (2014). The psychology of criminal conduct. Routledge.

Andrews, D. A., Bonta, J. (1995). Level of Service Inventory–Revised (LSI–R). Toronto, Canada: MultiHealth Systems.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). The Level of Service/Case Management Inventory (LS/CMI). Toronto, Canada: Multi-Health Systems

Barnes, G. C., & Hyatt, J. M. (2012). Classifying adult probationers by forecasting future offending. National Institute of Justice. Retrieved February, 4, 2020.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009). Forecasting murder within a population of probationers and parolees: A high stakes application of statistical learning. Journal of the Royal Statistical Society A, 172, 191-211.

Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. Criminal Justice and behavior, 36(1), 21-40.

Camilleri, H., Ashurst, C., Jaisankar, N., Weller, A., & Zilka, M. (2023, October). Media coverage of predictive policing: Bias, police engagement, and the future of transparency. In Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-19).

Craik, A., Han, L., Sullivan, L., Landsiedel, J., Travers, T., Spaull, C., Howard, P. (2024) Revalidation: Risk of recidivism tools. An evaluation of the actuarial instruments developed to assess recidivism risk in England and Wales. Ministry of Justice Analytical Series.

Comité Instrumentarium Forensische Psychiatrie. (2000). Risicotaxatie in de forensische psychiatrie: een Nederlands instrument in ontwikkeling. [Risk assessment in forensic psychiatry: Development of a Dutch instrument]. The Hague, The Netherlands: Dutch Ministry of Justice.

Dalal, D. K., Sassaman, L., & Zhu, X. S. (2020). The impact of nondiagnostic information on selection decision making: A cautionary note and mitigation strategies. Personnel Assessment and Decisions, 6(2), 54–64.

de Vries Robbé, M., De Vogel, V., & Stam, J. (2012). Protective factors for violence risk: The value for clinical practice. Psychology, 3, 1259– 1263.

Dobash, R. P., Dobash, R. E., Cavanagh, K., Smith, D., & Medina-Ariza, J. (2007). Onset of offending and life course among men convicted of murder. Homicide Studies, 11(4), 243–271.

Emeagi, C., Sullivan, L., Landsiedel, J., Craik, A. & Howard, P. (2024). The Actuarial Prediction of Sexual Reoffending Responding to Changing Offending Patterns. Ministry of Justice Analytical Series.

Eriksson, L., Mazerolle, P., Wortley, R., Johnson, H., & McPhedran, S. (2019). The offending histories of homicide offenders: Are men who kill intimate partners distinct from men who kill other men? Psychology of Violence, 9(4), 471–480.

Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. Journal of Criminal Justice, 81, 1–9.

Fazel, S., Wolf, A., Larsson, H., Lichtenstein, P., Mallett, S., & Fanshawe, T. R. (2017). Identification of low risk of violent crime in severe mental illness with a clinical prediction tool (Oxford Mental Illness and Violence tool [OxMIV]): a derivation and validation study. The Lancet Psychiatry, 4(6), 461–468.

Hanson, R. K., & Thornton, D. (1999). Static-99: Improving actuarial risk assessments for sex offenders (User Report 99-02). Ontario, Canada: Department of the Solicitor General of Canada.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. Criminal justice and behavior, 20(4), 315-335.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). Violent offenders: Appraising and managing risk (3rd ed.). American Psychological Association.

Havard, T., Nnamokon, N., Magill, C., Demeocq, C., Procter, J., Harvey, D., & Bettinson, V. (2023). Using Artificial Intelligence to Identify Perpetrators of Technology Facilitated Coercive Control.

Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.). Handbook of violence risk assessment, 11–28.

Herrera, F., Carmona, C.J., Gonzalez, P. & Del Jesus, M.J. (2011). An overview on subgroup discovery: Foundations and applications. Knowledge and Information Systems 29, 495-525

Home Office (2022), published 16 December 2022, https://www.gov.uk/government/publications/serious-violence-duty, Serious Violence Duty.

Home Office (2020), published 5 March 2020, https://www.gov.uk/government/publications/trends-and-drivers-of-homicide-main-findings, Trends and drivers of homicide: Main findings.

Home Office (2018), published 20 June 2018, https://www.gov.uk/government/publications/serious-violence-strategy, Serious Violence Strategy.

House of Commons Library (2023), published 11 July 2023, https://commonslibrary.parliament.uk/research-briefings/cbp-8224/, Homicide Statistics.

Howard, P., Craik, A., Han, L., Spaull, C. (2023) Escalation in the severity of offending behaviour. Ministry of Justice Analytical Series.

Howard, P. D., & Dixon, L. (2012). The construction and validation of the OASys Violence Predictor: Advancing violence risk assessment in the English and Welsh correctional services. Criminal Justice and Behavior, 39(3), 287-307.

Kropp, P. R., & Hart, S. D. (2015). SARA-V3: User manual for version 3 of the spousal assault risk assessment guide. ProActive Resolutions.

Latessa, E. J., Lemke, R., Makarios, M., & Smith, P. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). Fed. Probation, 74, 16–22.

Loeber, R., Pardini, D., Homish, D. L., Wei, E. H., Crawford, A. M., Farrington, D. P., Stouthamer-Loeber, M., Creemers, J., Koehler, S. A., & Rosenfeld, R. (2005). The prediction of violence and homicide in young men. Journal of consulting and clinical psychology, 73(6), 1074–1088.

Ministry of Justice (2025), published 24 April 2025, https://www.gov.uk/government/statistics/proven-reoffending-statistics-april-to-june-2023, Proven reoffending statistics: April to June 2023.

Ministry of Justice (2024), published 31 October 2024, https://www.gov.uk/government/statistics/proven-reoffending-statistics-october-to-december-2022/serious-further-offences-annual, Serious further offences annual.

Ministry of Justice (2022), published 8 April 2022, https://www.gov.uk/government/publications/ministry-of-justice-digital-strategy-2025/ ministry-of-justice-digital-strategy-2025, Digital Strategy 2025.

Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. International Journal of Forecasting, 30(3), 491-497.

Moore, R. (2015). A compendium of research and analysis on the Offender Assessment System (OASys). Ministry of Justice Analytical Series.

National Police Chiefs Council (NPCC) (2023), published 31 March 2023, https://www.npcc.police.uk/publications/policing-vision-2030, Policing Vision 2030.

National Council on Crime and Delinquency (NCCD) (2018). “System Manual Correctional Assessment and Intervention System.” National Council on Crime and Delinquency.

Nuffield, J. (1982). Parole Decision-Making in Canada: Research towards Decision Guidelines. Solicitor General of Canada.

Office for National Statistics (2023). 2021 Census: Aggregate Data. [data collection]. UK Data Service. SN: 8964, http://doi.org/10.5257/census/aggregate-2021-1.

Office for National Statistics (2025), released 6 February 2025, https://www.ons.gov.uk, Homicide in England and Wales: year ending March 2024.

Olver, M. E., Mundt, J. C., Thornton, D., Beggs Christofferson, S. M., Kingston, D. A., Sowden, J. N., Nicholaichuk, T. P., Gordon, A., & Wong, S. C. P. (2018). Using the Violence Risk Scale-Sexual Offense Version in sexual violence risk assessments: Updated risk categories and recidivism estimates from a multisite sample of treated sexual offenders. Psychological Assessment, 30, 941-955.

Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality. Information & communications technology law, 27(2), 223-250.

Richards, L. (2009). Domestic abuse, stalking and harassment and honour based violence (DASH, 2009) risk identification and assessment and management model. Association of Police Officers (ACPO).

Soothill., K., Francis, B., Ackerley, E., & Fligelstone, R. (2002). Murder and Serious Sexual Assault. What criminal histories can reveal about future serious offending. Police Research Series Paper 144. London: Home Office.

Soothill, K., Francis, B., & Liu, J. (2008). Does serious offending lead to homicide? Exploring the interrelationships and sequencing of serious crime. British Journal of Criminology, 48, 522–537.

Urbaniok, F. (2004). FOTRES: Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System (1st ed.). Bern, Switzerland: Zytglogge.

Viljoen, J. L., Goossens, I., Monjazeb, S., Cochrane, D. M., Vargen, L. M., Jonnson, M. R., Blanchard, A. J. E., Shanna, M. Y. L., & Jackson, J. R. (2024). Are risk assessment tools more accurate than unstructured judgments in predicting violent, any, and sexual offending? A meta-analysis of direct comparison studies. Behavioral Sciences & the Law, 43, 75-113.

Ward, T., Gannon, T., & Vess, J. (2009). Human rights, ethical principles, and standards in forensic psychology. International Journal of Offender Therapy and Comparative Criminology, 53(2), 126-144.

Webster, C., Douglas, K. S., Eaves, D., & Hart, S. (1997). Assessing risk for violence. Simon Fraser University, 251-277.

Wertz, M., Schobel, S., Schiltz, K., & Rettenberger, M. (2023). A comparison of the predictive accuracy of structured and unstructured risk assessment methods for the prediction of recidivism in individuals convicted of sexual and violent offense. Psychological assessment, 35(2), 152.

Wolff, K. T., & Baglivio, M. T. (2017). Adverse childhood experiences, negative emotionality, and pathways to juvenile recidivism. Crime & Delinquency, 63(12), 1495-1521.

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the violence risk scale. A treatment-friendly violence risk assessment tool. Psychology, Public Policy, and Law, 12(3), 279–309.

Yu, R., Molero, Y., Lichtenstein, P., Larsson, H., Prescott-Mayling, L., Howard, L. M., & Fazel, S. (2023). Development and validation of a prediction tool for reoffending risk in domestic violence. JAMA network open, 6(7), e2325494-e2325494.

Appendix A
List of "serious non-sexual violence" offences

This Appendix lists the offences that are considered "serious non-sexual violence" in this project. Where the term "serious violence" is used, this also refers to only these offences.

Home Office Code	Delius Code	Offence description
100	2	Murder
101	3	Murder (incl abroad) of persons aged 1 year or over/Genocide
102	4	Murder (incl abroad) of infants under 1 year of age
200	5	Attempted murder (incl abroad)/genocide
302	8	Conspiring or soliciting etc, to commit murder (incl abroad)/genocide
303	9	Assisting offender by impeding his apprehension or prosecution in a case of murder/Concealing commission of genocide
304	10	Intentionally encouraging or assisting commission of murder
305	11	Encouraging or assisting in the commission of murder believing it will be committed
306	12	Encouraging or assisting in the commission of one or more offences of murder believing one or more will be committed
400	13	Manslaughter (Category)
401	14	Manslaughter (incl abroad)
402	15	Infanticide
403	16	Child destruction
405	18	Manslaughter on grounds of diminished responsibility
407	20	Causing or allowing the death of a child or vulnerable person (Dom Viol, Crime & Victims Act 2004)
410	23	Management of an organisation whose activities cause death by gross breach of duty of care
411	1500014508	Cause/allow a child/vulnerable adult to suffer serious physical harm (Dom Viol, Crime & Victims Act 2004)
500	24	Wounding and other acts endangering life
501	25	Shooting, wounding, etc, with intent to do grievous bodily harm etc, or to resist apprehension (S18)
502	26	Shooting at naval or revenue vessels
504	27	Attempting to choke, suffocate, etc, with intent to commit an indictable offence (garrotting)
505	28	Using chloroform, etc, to commit or assist in committing an indictable offence
506	29	Burning, maiming, etc, by explosion
507	30	Causing explosions or casting corrosive fluids with intent to do grievous bodily harm
509	32	Placing, etc, explosives in or near ships or buildings with intent to do bodily harm, etc

510	33	Endangering life or causing harm by administering poison
511	34	Causing danger by causing anything to be on the road, interfering with a vehicle or traffic equipment
513	35	Possession etc of explosives with intent to endanger life
514	36	Possession of firearms etc with intent to endanger life or injure property etc (Group I)
515	37	Possession of firearms etc with intent to endanger life or injure property etc (Group II)
516	38	Possession of firearms etc with intent to endanger life or injure property etc (Group III)
517	39	Using etc firearms or imitation firearms with intent to resist arrest etc (Group I)
518	40	Using etc firearms or imitation firearms with intent to resist arrest etc (Group II)
519	41	Using etc firearms or imitation firearms with intent to resist arrest etc (Group III)
520	42	Use etc. of chemical weapons
523	45	Weapons related acts overseas
524	46	Use of noxious substances or things to cause harm or intimidate
526	48	Endangering safety at aerodromes
527	49	Torture (CJA 1988)
801	70	Wounding or inflicting grievous bodily harm (inflicting bodily injury with or without weapon) (S20)
802	71	Administering poison with intent to injure or annoy
803	72	Setting spring guns, &c, to injure trespassers
805	74	Assaults on persons preserving wreck
877	1500017509	Intentional Strangulation (Serious Crime Act 2015)
3701	332	Causing death by aggravated vehicle taking
5601	434	Arson endangering life
5700	436	Criminal damage (including causing explosion) endangering life (excluding arson)
5712	439	Criminal damage endangering life (excluding arson) - railway
5713	440	Criminal damage endangering life (excluding arson) - aircraft
5714	441	Criminal damage endangering life (excluding arson) - motor vehicle
5715	442	Criminal damage endangering life (excluding arson) - dwelling
5716	443	Criminal damage endangering life (excluding arson) - other building
5717	444	(Criminal damage endangering life (excluding arson) - other
5722	445	Crim Dam endangering life (excluding arson) - railway
5728	446	Criminal damage endangering life (excluding arson) - telegraph
833	98	Racially aggravated malicious wounding or GBH (s.29(1)(a),(2) of Crime and Disorder Act 1998)
840	105	Religiously aggravated malicious wounding or GBH (s.29(1)(a),(2) of Crime and Disorder Act 1998)
846	111	Racially or religiously aggravated malicious wounding or GBH (s.29(1)(a),(2) of Crime and Disorder Act 1998)

Appendix B
Top 20 local police features

This Appendix includes a list of the top 20 local police variables that were identified as important by an EBM model. This is the same model that was used to compare performance between data sources in Section 5.4. Feature importance was measured using the average absolute SHAP values, which broke down predictions into additive contributions.

Variable description	Feature importance ranking
Has the person been involved in an incident in an area that is not their probation area?	1
Current age	2
Number of arrests for assault or wounding	3
Number of crimes where a weapon was used to cause injury	4
Number of years since last contact with police	5
Number of years since first crime	6
Number of years since last arrest	7
Age at last crime	8
Age at last incident	9
Number of crimes where a weapon was used as a threat	10
Has the person ever been arrested?	11
Age at last arrest	12
Age at first crime	13
Years since first arrest	14
Number of crimes where the main group was violence with injury	15
Age at first arrest	16
Number of crimes where the main group was violence against the person	17
Number of crimes where the person was a suspect	18
Number of distinct victims throughout entire criminal history	19
Number of arrests for possession of a weapon	20