Ethics Committee Briefing Note—20TH November 2019

Project Reference: National Data Analytics Solution – Most Serious Violence (MSV)

Purpose of data analysis:

The strategy document 'Policing Vision 2025'[1] 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 National Data Analytics Solution (NDAS) aims to become a centralised advanced analytics capability for UK policing. UK police forces have access to a vast amount of digital data, but arguably lack the technological capability to use it effectively.[2] By proving that advanced analytical methods can be applied to existing law enforcement datasets, it is hoped that actionable insights grounded in data might be used to guide local intervention efforts and support the cross-cutting outcomes that evolved from the reform strands within the Policing Vision 2025. Putting information at the heart of decision-making in policing by connecting existing datasets for new insights should inform risk assessment and resource prioritisation.

NDAS demonstrated this capability during its Foundation Phase, which ran from September to April 2019. Three high-priority use cases were run as a proof of concept: Most Serious Violence, Workforce Wellbeing and Modern Slavery. Two of these use cases, Most Serious Violence and Modern Slavery, are nearly ready for operationalisation.

The founding partners of the NDAS are: West Midlands Police; Warwickshire Police; West Mercia Police; West Yorkshire Police; Greater Manchester Police; Merseyside Police, the Metropolitan Police Service; the National Crime Agency; and Staffordshire Police.

This submission to the WMP Ethics Committee concerns the Most Serious Violence (MSV) use case. This use case looks to use advanced predictive analytics to identify indicators that lead to an individual committing their first serious violence offence with a gun or a knife. Through applying these indicators, the model can output a risk score for all nominals known to the police, which can be managed and utilised as supplementary intelligence. The MSV use case relies on data from West Midlands Police, West Yorkshire Police; Warwickshire Police; and West Mercia Police.

Source of analytical question / hypotheses to be examined:

This use case aims to predict which individual nominals, who are already known to the police, are likely to commit their first most serious violence offence in the next 24 months. This problem statement is derived from the initial request from the Home Office: to create analytical models to provide new insights for serious violence based on previous patterns of offending and victimisation.

Data to be used:

Level of analysis

☒ Individual

Individuals aggregated?

☒ Yes

[1] HM Government, Policing Vision 2025

[2] Alexander Babuta, 'Big Data & Policing: An Assessment of Law Enforcement Requirements, Expectations and Priorities' (2017) https://rusi.org/sites/default/files/201709_rusi_big_data_and_policing_babuta_web.pdf

- No

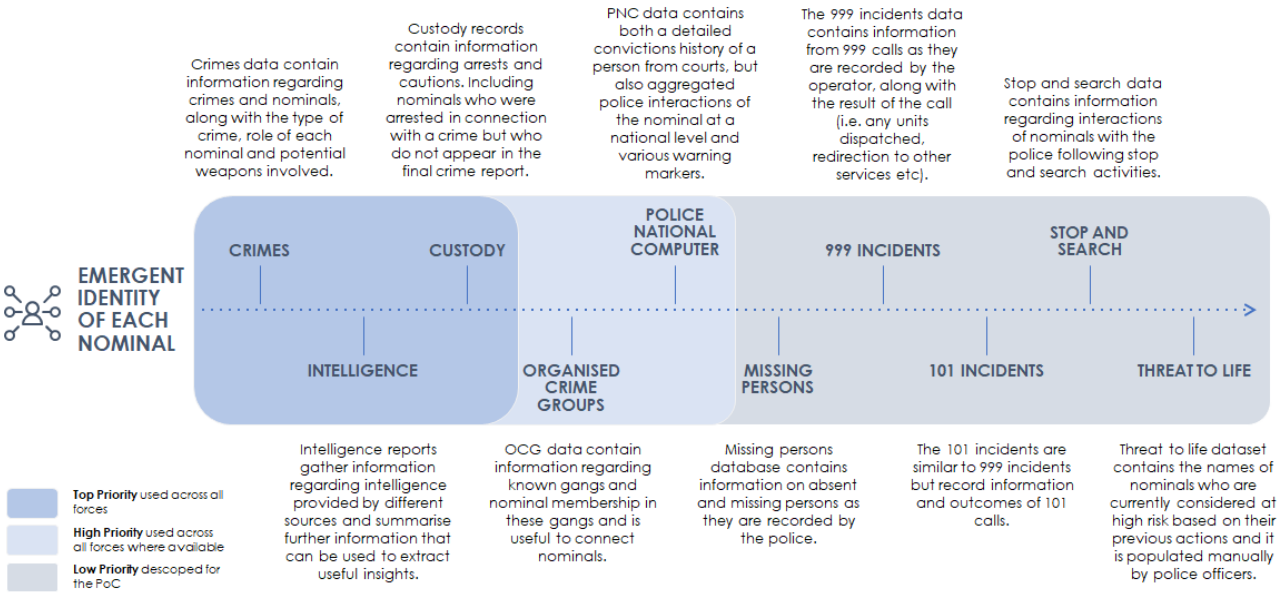
Specific Area:

- **Output Areas**
- **Super Output Areas - Lower**
- **Super Output Areas - Mid**
- **Wards**

Districts

West Midlands, West Yorkshire, West Mercia, Warwickshire

Multiple data sources were considered for the fulfilment of the MSV use case for the NDAS. 5 data sources (crimes data, intelligence reports, custody records, OCG and PNC data) were prioritised as the most significant to enable the development of the proof of concept model for MSV.



In this phase of work, this list consists of: Crimes, Intelligence, Custody, and Organised Crime Group (OCG) data.

Reliability of data:

The data is sourced from core systems used on a daily basis by the West Midlands Police, West Yorkshire Police, and Warwickshire-West Mercia Police. Once the data has been extracted, it is reviewed and verified by the NDAS team and any data quality issues identified are addressed with the guidance of data subject matter experts (SMEs). These discussions have enabled the identification of data quality issues and mitigations.

Before training the model, additional data suitability filters were applied. This included removing any variables with a high proportion of NULLs and those with very low variance. Exclusion of these variables improves predictive modelling capability through removing missing fields and variables that would have no effect on prediction.

Sample or entirety:

Entirety

If sample:

N/A

Method of sampling:

N/A

Method of choosing sample size:

N/A

Sample size:

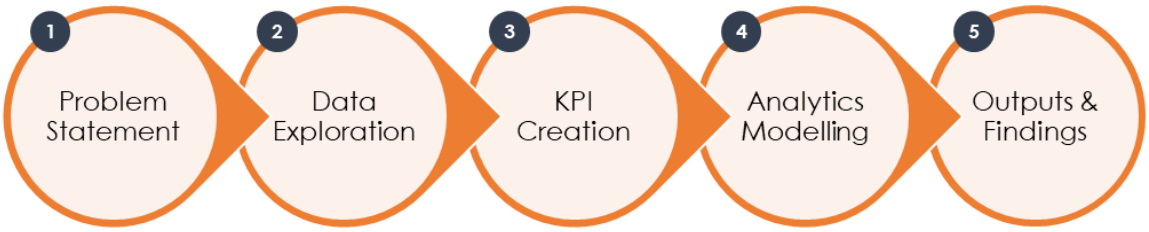
N/A

Type of analysis:

☐ **Exploratory**
☐ **Explanatory**
☐ **Predictive**
☐ **Optimisation**

Proposed methodology:

We are following a five-step process (outlined below):



1. Problem Statement

A fundamental input into the analytics process is the definition of the problem statement. The problem statement takes the initial request from the Home Office (to create analytical models to provide new insights for serious violence based on previous patterns of offending and victimisation) and builds out further detail about what is to be investigated. This allowed the NDAS team to define what should be termed as 'most serious violence' as well as identifying factors that precede an MSV offence that are significantly different from paths of offending that do not lead to an MSV offence. By adding more detail to the problem statement, the data points surrounding MSV can be more clearly identified.

A key note for the problem statement is that these are not the only crimes that will be investigated. The definition from the problem statement helps identify which crimes are classified as most serious violence and allows these crimes to be marked as the eventual 'aim' of the data science models, allowing us to identify patterns before the MSV event. The key predictive indicators identified in the proof of concept phase were identified in the context of the problem statement as defined in the initial phase of the NDAS project.

In that context the following conditions had to be satisfied to classify a most serious violence offence:

- The nominal had to have been charged with committing a crime of most serious violence with a gun or a knife
- The nominal cannot have committed a crime of most serious violence to date
- The prediction window is 24 months from the time the model is run

Having defined these conditions, our problem statement can be described as the ability to predict nominals at risk of committing their first MSV crime with a gun or a knife within a 24-month window from the time the model prediction is run. The problem statement as defined from the initial phase of the project will be reviewed and exact definition of the problem statement to be taken forward to production will be agreed with the partner forces involved.

2. Data Exploration

Once the problem statement has been defined, the NDAS team carry out a data exploration exercise. This involves investigating hundreds of tables across the five key data sources to find data fields that help to turn the problem statement from a business problem into a technical solution. The data exploration step helps to fully understand the available data that partner forces can provide to the National Data Analytics Solution.

3. KPI Creation

After the available data is understood, the data scientists will transform the raw data to create Key Predictive Indicators (KPIs). These Key Predictive Indicators can be calculated based on the available data to build a well-rounded viewpoint of an individual's timeline of appearances within police data. These KPIs will quantify behaviours that have either known or hypothesised associations with the MSV. KPIs will be created across three strands:

- Behavioural KPIs – these provide a summary of an individual's past behaviour derived from appearances within police data. Some examples of these behaviours are the past number of offences committed, the number of times a person has been a victim, or the number of times a person has been mentioned within an intelligence report. These are all based on past actual reported incidents that have been recorded on police systems, and do not use any information regarding race or location.
- Natural Language Processing (NLP) KPIs – these KPIs identify key data attributes from volumes of free text and extracts these attributes in a structured way. For example, translating bodies of text, which are difficult for a machine to read, into data attributes, such as, similarity to guns/knives represented as a number between zero and one. In this format, natural language processing can further distinguish and identify key differences in data items. This structured extraction of information can then be added to the behavioural information about a person.
- Network Analytics KPIs – by building a viewpoint of a person's network, it is possible to quickly identify how closely linked an individual is to known individuals charged with MSV crimes. MSV offenders. As this network changes over time, it is possible to see where people are potentially being drawn into networks of individuals who are charged with committing MSV offences.

Once the KPIs have been created to provide a rounded summary of known appearances within police data and associated individuals, these three sets of KPIs can be joined together to form an Analytical Base Table (ABT). This ABT can then be used as the starting point for analytical modelling.

4. Analytics Modelling

Once the data has been understood and brought together within the ABT, the analytical modelling process starts. The analytical models are created and trained on historic data available to police forces to help identify correlations between attributes that are indicative patterns or precursors leading to an MSV offence. These models are iteratively tuned to ensure that they are identifying the correct linkages and correlations. The results of the model are validated against unseen (out-of-time) data to ensure that the model is effective at identifying precursors of MSV.

5. Modelling Outputs

The final stage of the analytics process is to present the results of the model. These results are presented in a dashboard, which shows the findings from the MSV analytics model—both in terms of the overall strategic statistics of MSV within a police force area as well as providing a single view of an individual's past recorded appearances within police data.

Will the project eventually be automated:

☐ **Yes**
☐ **No**

Predictive models will always require re-training because predictive factors can change over time—the MSV analytical model will not be 'left alone' to process data. Continuous improvement will be built into the model monitoring process as the MSV analytical model will need attention and vigilance to ensure that the output generated is as accurate and unbiased as possible, and that any irregularities are addressed as soon as they arise.

In addition, as addressed below, the intention of the NDAS project is to continue to provide analytical insight to support decision-makers. A human decision-maker will have to form their own understanding of the factors giving rise to the output of the model before putting an intervention in place.

Means of evaluation:

Internal Technical Evaluation

The problem statement, methodology and Proof of Concept model results were reviewed as part of an internal evaluation to provide recommendations on how to improve the validity and accuracy of the analytical model. The reviewer was not involved with the NDAS project, and internal evaluations such as this are conducted as part of general quality assurance activities by the NDAS' delivery partner.

These were the recommendations from the internal evaluation:

- Training on a more balanced 50%-50% target variable distribution of 50% rather than 90%-10% was recommended; however, the precision score dropped significantly as more data was discarded due to under-sampling.
- Trialling different modelling approaches such as XGBoost, Support Vector Machine and documenting results in more detail. These approaches could influence the precision score.
- Using feature selection on tighter intervals to find the optimum number of variables to use as input to the model. This refines the variables to only the necessary variables to be used.

The NDAS team are working on the above recommendations and another technical evaluation will be planned in the next few weeks. Our own methodology/approach consists of regular model monitoring and capturing/reviewing the results that are generated as part of the output.

Independent Evaluation

In addition, it is critical that the analytical models and output developed by the NDAS are independently evaluated before moving into large-scale deployment as planned after this phase of work. Along with the College of Policing and WMP's Academic Research Lead, the NDAS is exploring options for partnering with an external organisation to provide an independent evaluation of both the MSV analytical model and its outputs.

The NDAS is committed to having the MSV model results be subject to independent academic evaluation. After the Foundation phase of work, we asked Dr Vicky Thakordas-Desai, a consultant forensic psychologist and HCPC registered psychologist, to conduct a review of the MSV model outputs (specifically the key predictive indicators) and assess whether these resonate with clinical experience and relevant academic research. Dr Thakordas-Desai will also be brought on during this phase of work as an independent

consultant to provide insight into forensic psychology and how a multi-systemic approach alongside an evidence-based risk assessment might enable at-risk young people and families to be identified and supported early on.

ALGO-CARE considerations:

Advisory

If applicable, are the outputs from the algorithm to be used in an advisory capacity?
Yes. The NDAS is not a tool that substitutes the professional judgment and discretion of law enforcement practitioners for automated decision-making. It is designed to assist human decision making without providing automated processing. It is intended that the outputs from the analytical model will help identify correlations between attributes that lead a person onto a journey of committing crimes of serious violence—the identification of these patterns could build a more informed, multi-systemic intervention approach. The output of the MSV analytical model will be provided as a supplementary source of information to assist officers and staff with their work, as opposed to stating an automated decision or recommendation for them.

Does a human officer retain decision-making discretion?
Yes. For the MSV analytical model to be used as a supplement to inform traditional policing strategies through established decision-making processes, meaningful review of the outputs generated will always be necessary. Outputs generated by the MSV analytical model will be used as supplementary information to support existing decision-making processes at the strategic, neighbourhood, or individual level, in conjunction with established statutory partnerships (Community Safety Partnerships and the West Midlands Violence Reduction Unit). These bring together experts from the police, Public Health England, the NHS Clinical Commissioning Group, the Local Education Authority, the Youth Offending Service and Probation to put in place evidence-based, multi-systemic interventions to support individuals, families and communities to be resilient to the factors that lead to violence within communities.

It is necessary to highlight the risk that output generated by an algorithmic system might be relied on too heavily in decision-making. In this context, this would amount to an illegal fettering of officer discretion—recognising this, the NDAS is working with the West Midlands Violence Reduction Unit as a test case to develop an approach for how MSV analytical output might be used to supplement decision-making.

We are working on the following:

1. Understanding the needs of our intended audience, including their process challenges and how MSV outputs (including data visualisations) will be designed to complement and improve the current process, while adhering to standards of usability, explainability, and transparency in mind
2. Building a supplementary toolkit to support decision-makers in interpreting and using the supplied output to plan and deliver interventions. This will consist of:
    - Governing principles to embed foundational values of fairness, accountability, and transparency to any decision-making framework where MSV outputs are used. Feedback from the Committee about this framework is welcomed, but it is suggested that the framework will be based on Floridi and Cowls' 'Unified framework of five principles of AI in society'[3]:
        - Beneficience
        - Non-malefecance

[3] Luciano Floridi and Josh Cowls (2019), A Unified Framework of Five Principles of AI in Society (2019), [URL: https://hdsr.mitpress.mit.edu/pub/l0jsh9d1], DOI: 10.1162/99608f92.8cd550d1.

- Autonomy
- Justice
- Explicability (intelligibility and accountability)

- Guidance to support intelligibility: plain language guidance to promote intelligibility of output, so that users can safely and effectively understand and interpret NDAS output to make assessments of risk in combination with their discretion and professional expertise. Any limitations of the output (e.g. as relates to false positives, accuracy, or data limitations) shall be communicated.
- Analysis of effectiveness of interventions: it is essential that data on interventions can be captured and fed into the MSV analytical model so that effectiveness (i.e. the impact of any interventions on the precision score) can be measured. This should not only refine the model output but allow decision-makers to understand the effect of interventions delivered.

Results of this partnership will be shared with others involved in the project, particularly the partner forces, and the Stakeholder Governance Group, which includes the Chair of the Ethics Committee, as an iterative process to determine the appropriate principles guiding the adoption of MSV outputs and what an 'interventions toolkit' might look like.

The NDAS welcomes suggestions from the Ethics Committee on other measures that might be undertaken as part of the approach outlined above. We intend to continue to engage with the Committee beyond operationalisation to ensure we can provide feedback on the results of this phase of work and adjust our approach accordingly.

Lawful

What is the policing purpose justifying the use of the algorithm (means and ends)?

The UK government's Serious Violence Strategy, published on 9 April 2018, is described as a 'very significant programme of work involving a range of Government Departments and partners' that emphasizes 'the importance of early intervention to tackle the root causes and provide young people with the skills and resilience to lead productive lives free from violence'. By supporting 'a new balance between prevention and effective law enforcement', the strategy underscores the importance that 'the police have the tools and support they need to tackle violent crime'.4

The first chapter of the strategy document outlines the government's analysis of the root causes of the rise in serious violence, from the influence of drugs profits[5] from county lines activity, through to adverse childhood experiences and vulnerability. A year on from the publication of this strategy, a House of Commons Home Affairs Select Committee provided an analysis of whether the Government's strategic response matches the scale of the rise in serious youth violence. Taking evidence from a range of sources including policing leaders, the Committee found that despite the government's strategy providing a 'relatively coherent analysis of the scale of the problem and the potential causes' and its 'commitment to a public health approach' in tackling serious violence, it argues that the Government's lack of a coordinated operational response is 'completely inadequate'. Further, the Committee argued:

> 'Although the strategy refers to risk factors for involvement in violence, its analysis is based largely on readily-available evidence. It is not underpinned by any attempt to collect data or gain a clear understanding of the number of people—particularly young people—at risk of serious violence. We fail to see how the Government can get a grip on this problem or pursue a public health approach without a clear understanding of the size and location of the

[4] HM Government, Serious Violence Strategy, 9 April 2018

populations most at risk, so that it can target resources effectively.'

The MSV use case in the Foundation phase was borne out of a request from the Home Office to create analytical models to provide new insights for serious violence based on previous patterns of offending and victimisation. This use case looks to use advanced analytics to identify predictive indicators that lead to someone committing their first most serious violence offence with a gun or a knife. Through understanding and applying these indicators, the model can output a risk score for all nominals already known to the police, which has the potential to be utilised as supplementary intelligence by the police. The NDAS presents a new opportunity to analyse data already held on police systems to refine the informational picture used in current strategic and tactical decision-making and will help forces to build an information-backed interventions approach.

The output generated by the MSV analytical model will provide digestible insights in the form of data visualisation dashboards to end users in partner forces, allowing them to drill into the detail of crime trends, risk levels, and KPIs that influence risk levels for nominals. As part of the current phase of work, we are working with the West Midlands Violence Reduction Unit to understand how insights generated by the MSV analytical model might help to support their approach to planning and delivering evidenced-based interventions.

Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes?

There is minimal impact to the privacy of individuals as the MSV use case relies on data that is already held on police systems.

We would argue that—because the potential interference to individual rights is not substantial; there is a logical connection using the MSV model for a legitimate policing aim (reducing violent crime); and that there is an appropriate existing system of oversight, challenge, and review—the MSV use case should be operationalised and the framework within which interventions are planned and delivered should continue to be developed. We welcome the experimental proportionality framework[6] that would permit the use of algorithms by public bodies such as the police while giving the public confidence that such tools would be used in a controlled way, which gives a limited presumption of proportionality to public bodies using algorithmic systems in decision-making, while allowing the NDAS to explore the benefits and outcomes as a result of operationalisation.

In what way will the tool improve the current system and is this demonstrable?

There is currently no formal system in place to undertake these types of analyses. There is currently no structured, scientific way to provide this level of insight into the circumstances and reasons that an individual might commit their first serious violence offence. The MSV output will assist the current formal process to identify risk and put interventions in place, allowing users to apply consistency in decision-making processes (which are potentially flawed due to the risk of bias) and to take full account of all data that is currently available. The value of the MSV use case will lie in pulling specific data points from its output that may help to build an understanding of the root patterns that lead to serious violence, particularly for young people.

Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim?

All data used by the NDAS is derived from existing police systems, meaning the data was obtained and processed for criminal law enforcement purposes under Section 3 of the Data Protection Act 2018. In

[6] Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018), Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality', Information & Communications Technology Law, 27:2, 223-250, DOI:10.1080/13600834.2018.1458455

addition, the acquisition, processing, and retention of data by the NDAS on behalf of West Midlands Police is governed by an information sharing agreement (ISA) between partner agencies. 4 partner agencies signed the ISA during the Foundation Phase. These include West Yorkshire Police, Warwickshire Police and West Mercia Police.

The ISA stipulates:

'Each national analytics assignment commissioned through NDAS governance will look to answer a specific problem (or "use case") on behalf of the Partner Forces, in line with one or more of the following policing purposes:

- Protecting life and property
- Preserving order
- Preventing the commission of offences
- Bringing offenders to justice, or
- Any duty or responsibility of the police arising from common or statute law.'

In this way, all data sources will be shared for a common, lawful and specified purpose.

In accordance with the Information Commissioner's guidelines, a full Data Protection Impact assessment was conducted for both the Foundation and Operationalisation phase. It is attached in the appendix to this submission. A legal review from the WMP legal team will follow in due course.

Is the operation of the tool compliant with national guidance?

There is currently no formal national guidance framework that governs the use of data analytics in law enforcement. The NDAS generally relied on the GDS Data Ethics Framework 2018, but it is increasingly clear that a robust regulatory framework is now necessary to govern this dynamic new application of technology in law enforcement. As highlighted in the Ethical Framework for the NDAS end-state (produced during the Foundation phase), the current combination of laws, codes, and practices relating to algorithmic systems in law enforcement will not drive the necessary legal and ethical approach to govern the rapid advancement of technology in this space.

The NDAS is part of the Home Office's Digital Policing Portfolio, which has articulated a national aim for a data analytics capability for UK law enforcement. In line with this aim, recommendations on national guidance establishing minimum standards on how data analytics platforms should be developed and used by law enforcement need to be produced. Considering its important role in shaping what these recommendations should look like, the NDAS welcomes this endeavour and will participate fully. In the interim, the NDAS will consider the independent research—commissioned by the CDEI and conducted by the RUSI—into the potential for bias to occur in predictive analytics technologies being developed by police forces.[7] The NDAS has been asked to participate in the Centre for Data Ethics and Innovation's effort to develop a code of practice for the trialling of predictive analytics technology in policing. [8] This builds on previous engagement undertaken by the NDAS with a range of stakeholders in navigating this landscape, including the Home Office Biometrics Commission; the National Police Chiefs' Council; the Royal United

[7] Centre for Data Ethics and Innovation, https://www.gov.uk/government/publications/report-commissioned-by-cdei-calls-for-measures-to-address-bias-in-police-use-of-data-analytics

[8] Centre for Data Ethics and Innovation Work Programme 2019 – 2020, https://www.gov.uk/government/publications/the-centre-for-data-ethics-and-innovation-cdei-2019-20-work-programme

Services Institute; the National Law Enforcement Data Programme; and the National Policing Information Risk Management Team.

In the absence of a framework regulating analytics in law enforcement, the NDAS has looked to ensure that its general operation remains aligned to the relevant existing national guidance that applies to law enforcement, particularly with regard to relevant data protection and administrative laws. Building on our completion of a Data Privacy Impact Assessment, NDAS is engaged in the Office for the Information Commissioner's Project DALE (Data Analytics in Law Enforcement) and is committed to continuing to ensure that all operations adhere fully with general data protection requirements for law enforcement.

Granularity

Does the algorithm make suggestions at a sufficient level of detail given its purpose and the nature of the data processed?

The proposed method produces suggestions and varying levels of detail which will be accessible and restricted (according to access management policies defined by working with partner forces and information assurance teams) to varying audiences based on the need for the data, meaning that sufficient levels of detail will be provided to only the appropriate colleagues. For example, some users would only need to analyse trends in data while others will need to understand the history and view of individuals, all of which is aligned and in adherence with the purpose of data processing.

Are data categorised to avoid broad-brush grouping and results and therefore issues of potential bias?

The underlying data is at an individual level and therefore groupings and categorisation are not applied by the solution, further analysis into groups will only be applied by the Officers.

Do the potential benefits outweigh any data quality uncertainties or gaps?

Yes. Firstly, no significant data quality issues or gaps were identified. Moreover, there is the social benefit of being able to identify the circumstances and reasons (particularly for young people) that lead individuals to commit their first violent offence with a knife or a gun, and subsequently using this insight to develop an informed interventions approach that seeks to prevent violent crime and young people from becoming criminalised.

Is the provenance and quality of the data sufficiently sound?

The data required comes from core source systems used on a day to day basis by forces and do not show evidence of significant data quality issues.

If applicable, how often are the data to be refreshed?

Currently data is planned to be refreshed weekly, but this is currently being reviewed as it has to be aligned with force IT departments.

If the tool takes a precautionary approach in setting trade-offs, what are the justifications for the approach taken?

The NDAS team have set a relatively high precision score cut-off (at least 50%) for the list of predicted nominals. This is to ensure we are minimising the amount of activity a force has to undertake by, in effect, minimising the number of false positives in the predicted cohort. We have chosen to minimise false positives by increasing the precision score cut-off so that we can generate a higher proportion of true positives, and the key predictive indicators that lead these to be identified by the MSV model.

Ownership

Who owns the algorithm and the data analysed?
As the lead force, WMP owns all models developed as part of the National Data Analytics Solution on behalf of the Home Office. Each respective force owns their own data being analysed and insights derived.

Does WMP need rights to access, use and amend the source code and data?
No.

Are there any contractual or other restrictions which might limit accountability or evaluation?
No.

How is the operation of the algorithm kept secure?
Partner Forces will be given a choice of SFTP (Secure File Transfer Protocol) or a Police Encrypted Device as a data transfer mechanism. Once received data will be uploaded to a WMP on-premise staging environment and then transferred to the secure NDAS AWS cloud, where all data will be held and processed throughout the delivery of the use cases. More detail on the security specifications can be found in the attached Data Protection Impact Assessment. This cloud solution will also go through a penetration test (estimated date November 2019).

Challengeable

What are the post-implementation oversight and audit mechanisms, e.g. to identify any bias?
A range of governance mechanisms are in place to monitor the progress of the NDAS project as it moves toward, and beyond, its first operationalisation phase:

- Strategic oversight is provided by the Home Office, as the NDAS is part of its Digital Policing Portfolio
- Tactical oversight is currently provided by the NDAS Stakeholder Governance Group, which includes the Chair of the West Midlands Ethics Committee
- Funding for the project is overseen by the West Midlands Office of the Police and Crime Commissioner
- The NDAS is engaged with the development of an ethical framework for the use of algorithmic systems in law enforcement, overseen by the CDEI
- The NDAS is engaged with the Office for the Information Commissioner on developing guidance over the use of data analytics in law enforcement

In addition, algorithmic systems deployed in this context need close supervision using ongoing human monitoring and auditing of performance against metrics such as accountability, bias, and security. As part of ongoing model maintenance and monitoring, the NDAS team will conduct regular reviews of the model to monitor against bias, including providing a statistical analysis of the impact of errors in the output (e.g. false positives and false negatives). The NDAS will also work with partner forces to support the development of a decision-making oversight and audit mechanism.

If the algorithm is to inform criminal justice disposals, how are individuals notified of its use?
The insights generated by the MSV analytical model will not be used to inform criminal justice disposals, including decisions on charge or bail; or decisions as to whether to continue an investigation into allegations concerning a subject. There will, however, be a process for notification, challenge and complaint within existing systems. If an intervention, guided by MSV output, is applied to an individual, they will be informed as to why the intervention is proposed. This will be in ordinary language and will provide the reasons behind why the individual has been identified for intervention. The MSV output will never be the

sole reason that an individual is identified for intervention, and therefore it will be explained in this light to the individual.

In addition to notification, external observers and data subjects shall be able to challenge the process by which an outcome was reached, to 'ensure that such tools are being used in accordance with the requirements of the relevant data protection legislation and principles of accessibility and natural justice under the Human Rights Act 1998'.[9] In line with this aim, NDAS governance forums should work with partner forces to embed standards for fairness, accountability, and transparency not just in the analytical models developed but also in the overall decision-making process that uses the outputs generated. For example, if a data subject wishes to challenge a decision that has been made with NDAS output (providing supplementary information to the decision-maker), a process will be established to allow the subject to scrutinise the model outputs, the key predictive indicators, and how these KPIs influenced the output.

Accuracy

Does the specification of the algorithm match the policing aim and decision policy?

Yes. The MSV use case was developed directly as a response to the problem statement which was defined in collaboration with partner forces.

Can the accuracy of the algorithm be validated periodically?

Yes. During the Foundation phase, data prior to 2014 was used as the training set and 2015-2016 data as the out-of-time validation set. The trained model made predictions on the out-of-time validation set using only the data trained prior to 2014.

Can the percentage of false positive/negatives be justified?

The Precision (Positive Predictive Value) score is the primary metric used to analyse the performance of the model—i.e. how many of the high-risk nominals predicted by the model actually committed an MSV over the 24-month period. It is calculated through dividing True Positives by True Positive plus False Positives. The Proof of Concept model achieved a 54% precision score for West Midlands Police and 74% for West Yorkshire Police on the out-of-time validation dataset. Crime events cannot be predicted with full accuracy (either due to interventions, impulsive behaviour, other random factors that may not be accounted for); therefore we believe that percentage of false positives can be justified.

The NDAS team are conducting ongoing work to fully understand the rate and statistical impact of false positives. While this work is still underway, where there are indicators to suggest that there are individuals linked to most serious violence within or just after the 24-month period, these results will be fed back and used to improve the accuracy and efficacy of the model in the future.

How was the method chosen as opposed to other available methods?

Logistic Regression, Gradient Boosting Machines and Random Forest algorithms were used for model training as these are the more prominent classification algorithms. Random Forest achieved higher performance in terms of precision score. Additional modelling approaches will also be tried with different parameters before operationalisation such as XGBoost and Support Vector Machines.

What are the (potential) consequences of inaccurate forecasts?

[9] Alexander Babuta, Marion Oswald and Christine Rinik, 'Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges' (2018) https://rusi.org/sites/default/files/201809_whr_3-18_machine_learning_algorithms.pdf.pdf

As stated above, the output generated during the MSV use case will be provided as a supplementary source of information for assisting officers and staff with their work, as opposed to stating an automated decision or recommendation for them. In combination with a supplementary toolkit to support how output should be used in decision-making practice, this means that the potential consequence of an inaccurate model result will be misallocation of resources (if the output is used to plan resources for interventions). Realistically inaccurate forecasts will have minimal impact on individual rights.

Does this represent an acceptable risk?
Yes. A false positive and false negative analysis will be conducted before outputs are provided to the end users at NDAS partner forces to mitigate the potential impact of output inaccuracy. As discussed above, this will be accompanied with a plain-language explanation of how the output was generated and the key predictive indicators that influenced the output. In addition, it is intended that police end users review and interpret the model's results—in combination with the associated key predictive indicators—to complement other sources of information in order to develop a targeted, well-informed interventions approach.

How are the results checked for accuracy and how is historic accuracy fed back into the algorithm for the future?
Data will be gathered for nominals at medium risk of committing an MSV crime provided by the model in order to check the accuracy of the model. As discussed previously, the model can be re-trained using the new data to mitigate model drift and performance can be improved over time with additional labelled data. In the future the NDAS team will continue to periodically evaluate the performance of the model, monitor and record the results so that the MSV model can be refined.

How would inaccurate or out-of-date data affect the result?
Data engineering pipelines are built to make sure erroneous or out-of-date data is not ingested. There will be strict data controls employed via production pipelines where all new data will be checked for any errors due to format. Dates of the records will also be checked to make sure data ingested is not out of date. Moreover, model prediction accuracy will be monitored closely in regular time intervals to track performance.

Responsible

Would the operation of the algorithm be considered fair?
One of the risks of deploying predictive analytics technology is the amplification and reinforcement of existing human biases. Some of these biases are unintended and arise from a lack of diverse perspectives when developing and training an analytical model. In other cases, decision-making can be skewed by reliance on incomplete data where other relevant factors are omitted. Finally, the historical data on which analytical models are trained may be biased. An analytical model merely manipulates data in order to produce an outcome. If the source data itself is biased, the model will inadvertently reflect biases inherent in the dataset.

In a law enforcement context, the perpetuation of these biases can be especially pernicious: for example, there has been much discussion over the potential for indirect racial bias to creep into analytical models used in law enforcement. It is well-documented that postcode information can function as a proxy variable for ethnicity or social deprivation, resulting in algorithmic outcomes that perpetuate bias.

Recognising these risks, ethnicity and gender variables were not used as inputs to the MSV analytical model. Similarly, location variables were also omitted to avoid this information becoming a proxy for other

factors, helping to mitigate any undue bias that may arise in the analytical model. The NDAS team is monitoring any bias in the data that might affect the identification of subjects by the model—this analysis will compare the characteristics of predicted nominals against the general population to validate the model's performance on fairness.

Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review?

Public engagement for transparency and accountability

Ensuring there is public trust in policing is paramount. The risks of damaging public confidence and trust in law enforcement are manifold, and the consequences well-documented. The public may be less likely to support the police in the pursuit of crime prevention, their willingness to participate as witnesses or to come forward as victims weakened. The application of advanced analytics adds complexity to this: although such technologies have been broadly applied in the private sector, it is still relatively new within policing—however, in both cases the level of public discourse is nascent. Compounding this, the opacity of how personal data is collected and used has understandably raised surveillance and privacy concerns—in addition, the pursuit of new technological initiatives without public consultation has arguably led to diminishing public trust in technology.

Despite these challenges, there is an opportunity for the NDAS to engage citizens to influence its operation and build local accountability in developing ethical approaches to the use of analytics in law enforcement. A plan for meaningful public engagement should play a role in this. The Royal Society of the Arts (RSA)[10] suggests the application of 'a process of citizen deliberation' in the deployment of analytics across three phases:

1. Public scrutiny through consultation when such systems are being introduced
2. Technical oversight through testing predictions for accuracy or expert-led auditing; and
3. Monitoring how the system is used by humans and evaluating it for accuracy

This submission has described plans for the NDAS to carry out points 2 and 3. We suggest that point 1, public scrutiny through consultation, be delivered through the Office of the Police and Crime Commissioner (at this stage, by the West Midlands OPCC). It is recognised that it takes real resource and commitment to deliver a plan for public engagement on this topic, and that consideration must be made towards being as representative of local citizens as far as possible to build local-level democratic accountability. Additionally, it is suggested that victim representation should be ensured within these groups and membership should be rotated regularly.

On a narrower level with regard to how this might apply to the MSV use case, it is important that public engagement be built into the process to an extent: for example, where output from the use case has been used to inform the Violence Reduction Unit's delivery of a targeted intervention on an individual, the individual must be fully informed of the overall decision-making process, allowing the individual to understand the output used, the key predictive indicators, and how these KPIs influenced the output. The NDAS will look to define what this process will look like with end users, supporting by providing plain language explanations of every stage of output generation.

Would it be considered to be used in the public interest and to be ethical?

Yes. To support police and partners in the development and delivery of multi-systemic, targeted

[10] RSA, 'Artificial Intelligence: Real Public Engagement' https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf

interventions to reduce serious violence and prevent people from criminalisation, it is arguably within the public interest for the MSV use case to be trialled operationally, in a way that is proportionate to any potential impact on individual rights.

Explainable

Is information available about the algorithm/decision-making rules and the impact of each feature?
Overarching standards of algorithmic transparency at every stage of the analytical process, as well as efforts to provide intelligible output, will support police decision-making. These are discussed above as part of our planned partnership with the WMP VRU.

A key driving factor for selecting an appropriate predictive model for the NDAS is interpretability and transparency. Trialling different data science techniques and models therefore allowed our data scientists to test whether similar results could be observed elsewhere and whether these techniques can be trusted. Three modelling approaches were used to understand which model could sufficiently capture the idiosyncrasies of the data. These are:

1. Gradient Boosting Machines
2. Random Forest
3. Logistic Regression

The Random Forest algorithm was chosen as the main classifier as it can provide the importance of each feature in terms of predictive power. The predictive variables can provide insight into the factors that may contribute to an individual committing a crime of most serious violence.

In addition to algorithmic explainability, the NDAS will aim to provide meaningful transparency to its end users by providing plain language information on the impact of each KPI in the output of the MSV analytical model. Linked to the common law duty to give reasons for significant decisions, wherever the MSV analytical output is used to supplement the decision-making process, the model and its output will be explicable and explainable to any individual upon which interventions are applied.