Ethics Committee Briefing Note

Project Reference: DAL_2019_0001_RASSO_analysis
Purpose of data analysis: To assess what makes for a successful investigation of RASSO (Rape and Serious Sexual Offences). In the face of increased reporting of these types of offences and static resourcing, the clear up rate resulting from these offences has been declining from circa 7.4% in April 2017 to circa 4.3% in April this year. There have also been issues raised about investigating these crimes at the national level, particularly in terms of how much digital information has to be examined in some circumstances and in other circumstances whether this information should be required. Because of these pressures identifying the elements of investigations that may improve the chances of a successful conclusion would allow improved resource use within WMP and potentially improved outcomes for victims. Please note that due to the nature of the project, any findings should NOT be made public in the interests of not providing information to perpetrators that may otherwise enable them to inhibit the progress of investigations to the detriment of victims.
Source of analytical question / hypotheses to be examined: The business question was posed by the FET.

Data to be used:
 	
Level of analysis:
 	
☐ Individual
   Individuals aggregated?
   ☐ Yes
   ☐ No
☐ Specific Area:
   ☐ Output Areas
   ☐ Super Output Areas - Lower
   ☐ Super Output Areas - Mid
   ☐ Wards
   ☐ Districts
☐ West Midlands
☑ Other
 	
The unit of analysis will be crimes.

Reliability of data: An extensive exploratory data analysis (EDA) phase will be undertaken to examine the extent of any data quality issues.

Sample or entirety:

Not applicable.

If sample:

Method of sampling:

Method of choosing sample size:

Sample size:

Type of analysis:

☑ Exploratory
☑ Explanatory
☐ Predictive
☐ Optimisation

Proposed methodology:

The project will involve an extensive EDA element, not only in terms of data quality issues, but also in terms of the relationship between various aspects of investigations and successful conclusions to them as well as the presence of any biases, etc. This is likely to involve an iterative process between the analytical base table (ABT) generation and EDA as well as EDA on the final ABT.

Following construction of the requisite definitions and ABT generation, at present it is considered that an initial explanatory model will be built which is likely to be a form of logistic regression model (which will enable an assessment of the strength of the relationship between aspects of an investigation and its successful or otherwise conclusion). The utility of building a directed acyclic graph (DAG) will also be examined (this form of modelling can be used to examine causal relationships).

Will the project eventually be automated:

☐ Yes
☑ No

Means of evaluation:

Not applicable (other than during the explanatory model build process).

ALGO-CARE considerations:

Advisory:

If applicable, are the outputs from the algorithm to be used in an advisory capacity?

The findings from the analysis may help identify the means by which RASSO investigations could be strengthened and as such may be used for future policy creation. It is not intended that this becomes a means by which investigations are determined to proceed or not (or other forms of prioritisation).

Does a human officer retain decision-making discretion?

Not applicable.

Lawful:

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

These types of investigations are, due to their nature, complex and potentially resource intensive. Changing technologies over the years has led to a change in the nature of investigations and, in some instances, the nature of the evidence that can be collected; this includes the increase seen in historic offences coming to light.

An ability to identify what leads to a 'successful' conclusion of a case and what detracts from success could therefore be of benefit for WMP in gaining an understanding as to what could be improved or mitigated in undertaking such investigations.

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

The analyses will involve data regarding RASSO, investigations and outcomes and therefore details personal to the victims of RASSO. However this needs to be balanced against the ability to potentially improve RASSO investigations which would be in line with the legitimate aims of WMP and which could aid help victims of such crimes in the future.

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

By providing an understanding of the factors that contribute to 'success', it is hoped that WMP could undertake RASSO investigations as effectively as is possible.

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?

The data to be analysed were collected as part of WMP's duty to investigate crimes and as such have been collected in the appropriate manner for the appropriate purposes.

Is the operation of the tool compliant with national guidance?

The analyses proposed would accord with DCMS Data Ethics Framework 2018.

Granularity:

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

The resulting analyses will provide information as to the effect 'on average' on the success or

otherwise of RASSO investigations. This is the appropriate level of detail to enable the identification of success factors in investigating these types of crimes.

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

Some categorisation may take place in forming the features of the explanatory modelling (e.g. broad age groups); however the details will become apparent during the analyses. Whilst the analyses will be on guard against the potential for paradoxical results that may arise from categorisation these are unlikely to be seen.

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

The project will include and extensive EDA element and this should highlight areas of heightened uncertainty in the data or where particular gaps exist. The benefits of these analyses, should they highlight areas within the control of WMP that could be acted upon, would be of benefit to victims of RASSO crimes and of course any resulting policy changes could be further examined.

Is the provenance and quality of the data sufficiently sound?

The data have been gathered during the day-to-day investigative work of WMP and do enable analyses of the type envisioned for this project.

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

Not applicable.

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

Not applicable.

Ownership:

Who owns the algorithm and the data analysed?

WMP owns the algorithm and the underlying data.

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?

The data and the analyses are contained wholly within the EMP Hadoop system and the security measures employed therein. The type of analyses do not lend themselves to on-going operation (in an automated fashion).

Challenge:

What are the post-implementation oversight and audit mechanisms, e.g. to identify any bias?

Not applicable

If the algorithm is to inform criminal justice disposals, how are individuals notified of its use?

Not applicable

Accuracy:

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

The nature of the analyses chosen have been determined to be the best means of addressing the research question.

Can the accuracy of the algorithm be validated periodically?

Not applicable.

Can the percentage of false positives / negatives be justified?

Not applicable.

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

Due to the nature of the research question (explanation as opposed to prediction). As noted above, the utility of using a DAG shall also be investigated.

What are the (potential) consequences of inaccurate forecasts?

Not applicable.

Does this represent an acceptable risk?

Not applicable.

How are the results checked for accuracy and how is historic accuracy fed back into the algorithm for the future?

Not applicable.

How would inaccurate or out-of-date data affect the result?

If data were to be wholly inaccurate then the analyses would essentially provide inapplicable findings. The Lab has sought to minimise this potential through a thorough analysis of the data in an extended EDA phase including their pitfalls, issues and overall nature. Discussions with SMEs should also highlight effects that may arise due to erroneous data.

Responsible:

Would the operation of the algorithm be considered fair?

The analyses will be fair in that each data point will be considered on its own merits.

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

The nature of the intended method(s) is such that the end-use is to provide information rather than predictions that feed into a process on an on-going basis.

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

In the face of increasing demand in this area it would be in the interest of victims if those elements that could improve the efficacy of investigations could be identified.

Explainable:

Is information available about the algorithm / decision-making rules and the impact of each feature?

The nature of the envisioned methods is such that information regarding the features and how they relate to each other is capable of explanation. HOWEVER, due to the nature of the analyses, the findings from the project should NOT be made public.