Written evidence submitted by Durham Constabulary (ALG0041)

Executive summary

- Durham Constabulary is exploring an algorithmic forecasting tool which identifies a suspect’s future risk of offending in order to provide decision support to police custody officers to reduce offending and minimise harm in communities;
- The algorithmic Harm Assessment Risk Tool (HART) utilises random forest machine learning methodology and forecasts three risk levels; high risk of committing a serious offence in the next two years; moderate risk of committing a non-serious offence in the next two years; and low risk, unlikely to commit any offences over the next two years;
- The random forest methodology has features which allow the balancing of different types of errors (dangerous and cautious);
- HART is able to forecast with 98% accuracy that for those forecast as low risk the very dangerous error will not occur. In order to do this the model will over predict cautious errors;
- Any variable can be omitted from a model, however, this may affect the accuracy of the model leading to more errors in forecasts. There is a balance between accuracy and fairness which in turn translates into an effect on communities if the model is used as decision support;
- Consideration should be given before the use of such models as to what the existing process is for the decision/risk area being considered;
- ‘Algorithms in Policing – Take Algo-care™’ is a proposed decision-making framework (see parallel submission) for the deployment of algorithmic assessment tools in policing and identifies some practical considerations for utilisation of tools such as HART;

Introduction

1. This submission describes an algorithmic forecasting tool which predicts future offending at the gateway to the Criminal Justice System (CJS), and is utilised in a police custody setting. Known as the Harm Assessment Risk Tool (HART), this tool is the result of a collaboration between the University of Cambridge and Durham Constabulary. The forecasting model supports the decision making of police custody officers by assessing the risk of future offending.

2. The tool was developed to identify offenders who are eligible for an intervention (Checkpoint) that is thought to reduce offending, which in turn will prevent future harm in communities. The submission also includes aspects of one author’s research regarding algorithmic forecasting in a criminal justice probation setting in the United States, which utilises the same algorithmic forecasting methodology.

3. HART was created as part of a Checkpoint programme, which is a culture changing initiative within Durham Constabulary. Checkpoint seeks to tackle the root causes of offending and associated health and community related issues. The programme identifies why an individual adult has offended, along with the best interventions and services to support the individual in turning away from crime. In order to pursue this strategy, Checkpoint must identify those at an appropriate risk of reoffending.
4. The forecasting tool allows the police to identify those who are most appropriate to receive a Checkpoint intervention, and do so at the initial police custody disposal decision point. Effective forecasting leads to effective triage, that targets appropriate offenders (based on risk of future offending) with the optimal intervention to, support a desistance from committing crime (Sherman, 2012).

5. In this submission, we do not comment on any potential or hypothetical uses of algorithms in policing. Having utilised this kind of model to predict general future offending, however, we can see numerous benefits in identifying individuals through this method, and using forecasts to prevent future criminal behaviour. We also suggest that other police forces will inevitably begin to explore similar methods in their risk assessments.

6. It is important, at this stage, to point out the forecasting model is being used to support the identification of cases suitable for Checkpoint. The Durham model does not yet provide custody officers with the full results of its forecasts, and is not yet used to support any decision-making beyond the enrolment of offenders into the Checkpoint programme.

Extent of current and future use of algorithms in decision-making

7. This submission describes an algorithmic forecasting tool which is currently utilised within the policing and criminal justice context. The tool supports decision making in the police custody environment, shortly after offenders have been arrested by the police and have reached the initial gateway to the criminal justice system. While HART forecasts support the custody officer’s decision making, they quite explicitly do not remove the officer’s discretion.

8. The HART algorithm was developed as part of an ongoing collaboration between Durham Constabulary and the University of Cambridge. The central goal of the development team was to promote consistency in decision making, enabling targeted interventions to rigorously test what works in preventing harm and reducing reoffending (Berk et al., 2009; Berk, 2012; Barnes and Hyatt, 2012; Neyroud, 2015; Sherman, 2012;)

9. The algorithm deployed in Durham was constructed using a form of machine learning known as random forests. This technique offers desirable features such as an ability to detect relatively rare but dangerous outcomes, to model relationships in non-linear ways, and to balance the differential costs of different kinds of errors (Barnes and Hyatt, 2012).

10. The current model produces forecasts which separate offenders into three different predicted risk groups. First, suspects who are predicted as likely to commit a new serious offence over the next two years are placed in the High Risk group. Secondly, those whose forecasted offending over this same time frame will be limited to non-serious crimes are designated as Moderate Risk. Finally, those who are predicted to commit no new offences during the next two years are identified as Low Risk. For these purposes, serious offences are defined, for example as, homicide, attempted homicide, highly violent offences (e.g. grievous bodily harm), robbery, sexual offences, and firearm offences.

11. The model is built using approximately 104,000 custody events over a five year period (2008-2012), and has been separately and independently validated using 15,000 custody events
from another full year (2013). It uses 34 different kinds of predictors to arrive at a forecast, the vast majority (29) of which focus upon the offender’s history of criminal behaviour.

12. The random forest model employed here uses 509 separate classification and regression decision trees (CART), which are then combined into the full forecasting model. Each individual tree is produced using both a randomly selected subset of cases and a randomly selected pattern of predictors. Essentially, each tree is a model in and of itself, and produces a forecast which is then used as one vote out of 509 total votes. The votes are counted, and the overall forecast for the full model becomes the outcome which receives the most votes (Breiman, 2001; Berk et al., 2009; Barnes and Hyatt 2012; Berk et al., 2016).

13. As with any forecasting effort, HART inevitably produces errors. In this case, however, with the random forests technique allows us to recognise different types of errors and the costs associated with them. The model built for Durham Constabulary recognises that all errors are not equally problematic. This method differentiates between an offender who is predicted to be relatively safe, but who then goes on to commit a serious violent offence (high risk false negative or very dangerous error), from an offender who is predicted to be at high risk of committing a serious offence, but who turns out to be low risk (high risk false positive or very cautious error). While both of these examples are errors in forecasting, the consequences and community impact are very different.

14. HART (along with its independent validation) demonstrates that, for those forecast as low risk the model can, with 98% accuracy, ensure that the very dangerous error described above will not occur. In essence, the more dangerous situations where risk is underestimated are given a higher cost within the model, which leads to them occurring less frequently. As a necessary consequence, however, this means that overestimates of risk are provided with a lower cost, and therefore become more likely to result from the model.

15. The data used as predictors in this algorithm are limited to those held within Durham Constabulary systems, and does not utilise data from other police force areas or indeed from national IT systems such as the Police National Computer or the Police National Database. This is one reason that such models could not currently function as the ultimate decision maker at the gateway to the CJS. The model simply does not have all of the information available to it, and can therefore only support human decision-making, rather than replace it. The custody officer may frequently be aware of other information that changes the risk level, and must apply their own judgement in deciding upon the disposition of each offender’s case.

The scope for algorithmic decision-making to eliminate, introduce or amplify biases or discrimination, and how any such bias can be detected and overcome.

16. As mentioned earlier, the Durham tool uses 34 predictors values, the majority of which (29) relate to the suspect’s offending history. These behavioural predictors are combined with age, gender, two forms of residential postcode, and the count of existing police intelligence
reports relating to the offender. Some of the predictors used in the model, therefore, relate to characteristics that offenders are unable to change, while others (such as postcode) could be viewed as indirectly related to measures of community deprivation.

17. In a model of this nature there is a clear need to understand how forecasts are arrived at and for serious consideration to be given to the ethics around using such models. It is the authors opinion that the purpose of the model should firstly be fully understood. In the case of Durham Constabulary, the key goal is to consistently identify offenders who present the proper risk level (Moderate Risk) to qualify for Checkpoint, while minimising the potential for harm in our communities. The algorithm allows the police to identify those suspects who are likely to reoffend and who could benefit from the Checkpoint intervention, while also identifying those suspects for whom it is safe to offer an out of court disposal.

18. There is an obvious concern with amplifying (or at least perpetuating) biases in models of this nature as the algorithms are built using historical outcome data. The model makes its predictions about the future based on offenders who have been arrested in the past. One could argue that, if the police target a deprived, high density neighbourhood/urban postcode, then more people from this area will come to police attention and be arrested than those living in an affluent, low density neighbourhood/rural postcode that does not see much police activity. The likelihood of arrest is therefore lower based on historical outcomes in the affluent, low density rural postcode or neighbourhood. It is the outcomes that build the model – as opposed to the model itself – that is of central concern. There are examples in the US media which suggest a disproportionate impact of forecasting across racial groups, leading to potentially uncomfortable consequences (Propublica article). Please also see the parallel submission relating to ‘Algorithms in Policing – Algo-Care’ and the legal/ethical framework developed in collaboration with Durham Constabulary and Winchester University.

19. There is, however, a level of reassurance in the fact advanced algorithms such as random forests are based upon millions of nested and conditionally-dependant decisions points, spread across 509 unique trees. Unlike earlier methods of forecasting, it is simply not the case that a given input on a single predictor has an inflexible and inescapable impact on the forecasted outcome. Simply residing in a given post code has no direct impact on the result, but must instead be combined with all of the other predictors in thousands of different ways before a final forecasted conclusion is reached. It is therefore the combination of variables, and not the variables in isolation, that produces the risk level (Barnes and Hyatt 2012; Berk, 2012).

20. In building these models, an organisation decides which variables are included in the model, balancing fairness and accuracy (Berk, 2016). Any variable can be omitted from the model, but doing so will invariably reduce accuracy. These are finely balanced decisions that need to be considered alongside the fundamental goals of consistency and preventing crime.

21. In Durham Constabulary, the forecasting model’s outputs are advisory, and it is the authors’ view that these models should not remove discretion from the police officer. No algorithm can have access to every piece of information that could be relevant to reoffending. When
the model does not have all of the information an officer can access, it cannot be the arbiter of any decision, including the decision to provide a suspect with an out of court disposal or pursue a court prosecution.

22. The model is built using the outcomes based on the historical reality reflected in police and criminal justice data systems, and this history is at least partly based on the human judgements that were made at the time. It would be wrong, and the error rates would increase, if the model failed to reflect reality. Before concluding that algorithms should therefore be viewed as biased, it is necessary to consider whether human judgement is more or less biased. Algorithms can usually estimate their own error rates, and it is tempting to compare them to some hypothetical perfection. Instead, however, it is better to compare them to the conditions that would persist if such models were not available. There are several points in the CJS where clinical discretion based decisions are made; arrest, bail, charge, court hearings, trial, and sentence. These chains of decisions produce a build-up of biases throughout the process, a forecasting model reflects that reality (Berk and Hyatt, 2015).

23. Durham Constabulary are also using the current model to understand how custody officers form their own subjective assessments of risk. This practice will eventually allow us to compare accuracy of the model’s predictions to those of experienced custody officers. No results are available at this time, however.

24. Understanding how the algorithm works, how its predictions compare to human judgement, and considering the finely balanced ethical issues are at the heart of Durham Constabulary’s desire to use this technology responsibly. It is in placing these matters at the heart of the work that has led Durham Constabulary to explore this area of work within an evidence based framework in collaboration with academia.

25. Algorithmic decision making and the complexity of machine learning means that it is becoming increasingly difficult to explain to non-computer scientists and non-statisticians how a machine learning forecasting tool arrives at its outcomes. It may not be appropriate for the detailed operation of a tool used in the investigative process to be ‘transparent’ to the public. The authors’ do recognise however, a need to make information available to an affected individual where a tool assists in the decision-making concerning an out-of-court disposal.

26. The authors submit that this objective could be fulfilled with a very basic notification in the custody environment, pointing to where fuller explanations regarding the purpose of the forecasting tool could be provided, for example, on a website.

27. It is important to resist the urge to compare the transparency of algorithmic forecasting to some mythical perfect transparency. Instead, it is useful to compare the potential
transparency of these models to the nearly-non-existent ability to gather similar measures regarding the everyday human decision makers. The biases of algorithms can, at the very least, be measured and presented in mathematical terms. The biases of human beings are far more inscrutable.

28. The Constabulary, in recognising the issues raised by utilising algorithmic forecasting has collaborated with Winchester University to develop a legal/ethical framework referred to as Algo-care (see parallel submission). This framework seeks, in a practical sense, to assist police forces considering the use of algorithmic forecasting.

29. Algo-care identifies that the police should be able to explain the decision-making rule(s) within such models, and provide the impact that each predictor has on their outcomes. The framework also suggests that agencies have access to and can deploy a data science expert to explain and justify the algorithmic tool (in a similar way to an expert forensic pathologist).

Regulatory oversight of algorithmic decision-making

30. The work performed by Durham Constabulary centres on the gateway to the CJS, very shortly after arrested offenders have been taken into custody. Decisions at this gateway are made by custody officers. In considering the perceived fairness of algorithms, it is therefore necessary to compare it to the fairness of clinically-biased decision making by these human decision-makers. Custody suites operate around the clock, with more than one individual custody officer rotating into the clinical decision-making role for each offender. For any given year of operation, those in custody will encounter a diversity of differently-biased judgements of not one but many custody officers, highlighting the need to assess risk in a consistent way. (Neyroud, 2015)

31. At various points in time, the police across the UK have been criticised for the current lottery of decision making at this point in the process. At the very least, models such as this offer some level of consistency in support of the decision at the gateway (Neyroud, 2015).

32. Algorithms such as HART also have the potential to reduce demand, not only for the police but for other CJS agencies. When demand reduction is discussed with the introduction of such models, critics can centre their attention on a distaste for such models, suggesting the aim is solely to achieve demand reduction and reduce costs. However, quite the opposite is true. The police will seek to prevent crime and harm by exploring new and innovative ways of ensuring resources work as efficiently as possible. By ensuring suspects of a particular risk level receive an intervention that promotes their desistance from crime, if successful, this can leverage this cessation of offending to minimise harm in communities. This effectively means communities are better supported by the police.

33. Within the policing environment, there are various points at which the police are assessing risk to predict where harm is likely to occur in order prevent crime and reduce harm. Similar random forest models, such as the Durham algorithm described above, could be utilised in other areas where risk is assessed. Examples of such areas include the most serious and demand-intensive areas of policing, such as sexual offender assessments, domestic abuse,
and cyber fraud. If we focus resources on those who are most likely to not only reoffend, but reoffend in a serious way, then we are adhering to the Peelian principles of preventing the worst amount of harm to society (Home Office, 2012).

34. Forecasting decision support is not a new concept, and various decision support structures have existed for decades. There must be an acceptance that, as time passes, more innovative modern decision support tools will become available.

35. In reducing crime and rehabilitating offenders to desist from crime, the police work in accordance with Peelian Principles to prevent crime, which in turn keeps victims and wider communities safe (Home Office, 2012). The prevention of crime and apprehension of the offender, together with their rehabilitation and conviction as identified by HMIC (2016), ‘are among the highest obligations of the state in the discharge of its duty to protect citizens’. It is therefore incumbent upon the police to ensure that, if a more modern method of effectively targeting citizens exists to minimise harm in the community, it should be fully explored within an evidence-based framework.

36. The authors submit that algorithmic models will continue to grow at a pace as police forces understand the capability of such models. To keep in line with the likely expansion of the use of algorithmic tools within policing, it is suggested that regulatory oversight and national guidance be developed by the College of Policing, overseen by HMIC.

April 2017

References


