

Written evidence submitted by Sheena Urwin, Head of Criminal Justice,  
Durham Constabulary (ADM0032)

I write further to my appearance before the committee on 12<sup>th</sup> December 2017 in relation to the inquiry into *Algorithms Used in Decision Making*. I am the Head of Criminal Justice in Durham Constabulary and following two submissions to the inquiry I was invited to appear before the committee. During the session on the 12<sup>th</sup> December I offered to write to the committee to provide data concerning how often the risk assessment by the police officer is different to the algorithmic assessment of risk (p.33 Q192, Q193, and Q194 refers).

Durham Constabulary has introduced a machine learning tool called HART – Harm Assessment Risk Tool to support decision making. The purpose of the model is to assist decision making relating to whether a suspect could be eligible for a deferred prosecution, known as Checkpoint, within Durham Constabulary. The purpose of the deferred prosecution is to understand which interventions work most effectively in encouraging offenders away from a life of crime.

Within the custody environment there are many different kinds of decisions that are made such as; police bail (conditional and unconditional); no further action being taken against a suspect; an out of court disposal for an offence committed by a suspect such as a deferred prosecution; charging a suspect (less serious decisions made by police whilst more serious decisions are made by CPS) and following charge, court bail (conditional or unconditional) or custody with the defendant being placed before the next available court. The custody officer must consider other factors in the decision making process (as outlined my evidence), the Harm Assessment Risk Tool (HART) is one factor of the many the custody officer considers, indeed is statutorily obliged to consider under both the Policing and Crime Act 2017 and the Bail Act 1976.

The purpose of the HART model is to assess the risk of reoffending over a two year period in a serious or non-serious way. The aim is to support the identification of those offenders for whom a deferred prosecution may be suitable. I outlined during my evidence there are many other factors custody officers consider as part of the decision making process besides the HART forecast. HART is part of a collaboration with the University of Cambridge and is part of ongoing research which will be published in due course, and is it for this reason I am only able to provide limited detail to you in answer to your questions.

HART assesses risk of reoffending and as such in order to test whether police officers assess risk in a similar way or disagree a process of *willful blindness* was used.

Essentially willful blindness entails the HART risk forecasts being conducted during the normal course of business in the background however not displaying the forecast risk level to the custody officers - the custody officers were unable to see the forecast risk. Whilst HART was willfully blind custody officers were asked in their judgement what the risk of reoffending was for the suspect before them. We are therefore in a position to understand how often the custody officer took a different view around risk of reoffending. The two tables below shows the agreement levels not accuracy levels.

Written evidence submitted by Sheena Urwin, Head of Criminal Justice,  
Durham Constabulary (ADM0032)

| Algorithmic vs Clinical | Police High | Police Moderate | Police Low | Model Total |
|-------------------------|-------------|-----------------|------------|-------------|
| Model High              | 1.58%       | 11.49%          | 2.03%      | 15.09%      |
| Model Mod               | 3.49%       | 39.86%          | 13.29%     | 56.64%      |
| Model Low               | 1.35%       | 12.16%          | 14.75%     | 28.27%      |
| Police Total            | 6.42%       | 63.51%          | 30.07%     | 100%        |

*Table 1 HART forecast and Custody Officer Agreement levels n=888:*

As can be seen in Table 1 above police officers forecast a high risk of offending in a serious way 6.42% of the time and HART forecasts high risk of offending in a serious way 15.09% of the time. Some of the increased high risk HART forecasts will be due to the way the model is deliberately built to overestimate in the high risk area to minimise the worse errors in the low risk area (as outlined in the written submissions to the committee). Police officers tend to forecast a moderate risk of non-serious offending 63.51% of the time whilst HART forecasts moderate risk of non-serious offending 56.64% of the time. Low risk of offending, the definition of which is no offending over a two year period, is very similar.

| Description  | Percentage |
|--|------------|
| Of those forecast by the model as High risk percent of occasions the police agreed     | 10.45%     |
| Of those forecast by the model as Moderate risk percent of occasions the police agreed | 70.38%     |
| Of those forecast by the model as Low risk percent of occasions the police agreed      | 52.19%     |
| Of those forecast by the police as High risk percent of occasions the model agreed     | 24.56%     |
| Of those forecast by the police as Moderate risk percent of occasions the model agreed | 62.77%     |
| Of those forecast by the police as Low risk percent of occasions the model agreed      | 49.06%     |

*Table 2: Agreement levels by risk n = 888*

Table 2 highlights specific forecasts and whether the model and police custody officer are in agreement. From the table the high risk area shows the highest levels of disagreement with low risk forecasts being the equivalent of flipping a coin.

Written evidence submitted by Sheena Urwin, Head of Criminal Justice,  
Durham Constabulary (ADM0032)

The data above demonstrates the levels of agreement however whether HART is more accurate in its forecast than the police officer remains to be seen. The forecasts are over a two year period and therefore we do need to wait for that time horizon to pass before we can say whether HART is more accurate and a good decision support tool to aid custody officer decision making.

It is important to emphasize two things, firstly that the data in the tables are not accuracy levels and purely relate to levels of agreement. Secondly, the forecast risk is not the deciding factor in the custody officer's decision, as outlined in my evidence, risk of reoffending is one factor of many. The custody officer makes finely balanced decisions which as I outlined on the day, take a number of factors into account besides risk of reoffending. I hope the above is of assistance and answers the questions the committee had regarding agreement levels.

If I may turn to a previous evidence session, to clarify some points made during your first session. On 14 November 2017 I was not in attendance at the session however did watch the evidence. It was suggested (p.22 Q34 of the transcript) by Professor Amoores that, 'They realise that a false negative could be a very risky thing, because that would mean not arresting somebody or keeping them in custody when they potentially had a high risk of reoffending. They talk about how they adjust the parameters in their algorithm, which means they tolerate a few more false positives—people who perhaps should not have been kept in custody for an extra 24 hours'. I have assumed we are talking about high risk false negatives/positives in this commentary. To be clear, the risk of reoffending does not determine whether or not someone is arrested, nor does it determine whether someone will be detained in custody for an extra 24hrs. HART is not determinative and we have legislation which governs and regulates these arena which would preclude such use.

Secondly at the conclusion of evidence to the committee on 14<sup>th</sup> November 2017 a question was posed regarding how often the custody officer had contradicted the HART decision. There was a suggestion (p.43 Q85 of the transcript) by Silkie Carlo from Liberty that a request to Durham Constabulary had been made asking that specific question and that Durham Constabulary had not provided the information – I can confirm no such request had been made. Had a request of that nature been made we would not have been in a position to provide it. HART does not make decisions but provides a forecast of risk of offending which is one factor of many the custody officer is statutorily obliged to consider, as outlined in my evidence. The tables detailed above are in the public domain and we are unable to publish any further detail as insufficient time has elapsed. It is our intention to publish research as and when it becomes available. I do appreciate there is a desire to see as much information as possible about HART, we too are keen to see the outcome of the research however we must wait for sufficient time to elapse.

I would like to thank the committee for the opportunity to submit evidence and affording me the opportunity to appear before the committee and suggest a knowledgeable regulatory body to provide oversight for the use of machine learning tools is necessary. Checkpoint and HART is a piece of research which is necessary to ensure we provide the very best decision support to our officers. The research is

Written evidence submitted by Sheena Urwin, Head of Criminal Justice,  
Durham Constabulary (ADM0032)

necessary to assist us to understand which interventions work best to support offenders away from a life of crime by addressing underlying issues. Essentially it is necessary to understand what works in policing through evidence based practice. It is our intention to publish the outcome of the research when sufficient time has elapsed both for the comparison between HART and the custody officers and the Checkpoint deferred prosecution which was recently highlighted in The Lammy Review, an independent review of the treatment of, and outcomes for, Black, Asian and Minority Ethnic (BAME) individuals in the Criminal Justice System (CJS).

*February 2018*