

Prison Reform Trust – Written evidence (NTL0004)

The Prison Reform Trust (PRT) is an independent UK charity working to create a just, humane and effective penal system. We do this by inquiring into the workings of the system; informing prisoners, staff and the wider public; and by influencing Parliament, government and officials towards reform. The Prison Reform Trust provides the secretariat to the All Party Parliamentary Penal Affairs Group and has an advice and information service for people in prison.

The Prison Reform Trust's main objectives are:

- reducing unnecessary imprisonment and promoting community solutions to crime
- improving treatment and conditions for prisoners and their families
- promote equality and human rights in the criminal justice system.

www.prisonreformtrust.org.uk

1. We are grateful for the opportunity to contribute to this inquiry. Our evidence concerns a single example of the use of an advanced algorithm in the categorisation process, a key aspect of prison management, which affects the daily experience of the great majority of prisoners in England and Wales.

2. For many decades, prisoners have been categorised in order to allocate them to what is seen as a prison with a suitable level of physical security. The process is designed to facilitate both an efficient use of the prison estate, and to prevent unnecessarily harsh restrictions on the freedoms prisoners can enjoy within the prison perimeter. For any prisoner, the prison in which they are held has a major influence on the quality of their life. For prisoners serving very long sentences, the promise of progression through security categories, from the most secure eventually to "open" conditions where they will also have the possibility of spending time in the community on temporary release, is of huge significance. For those whose release is dependent on a judgement about the risk they may pose to the public, that progression is vital to their chances of securing a favourable decision.

3. Decisions about categorisation carry great weight for the prisoner and for their family and loved ones. But they also have significance for the public and for ministers. A poor decision might lead to an avoidable escape, or a crime committed while the prisoner is on temporary release. A small minority of prisoners may use the relative freedoms of lower security prisons to continue to commit crime and undermine the safety of that prison for those around them.

4. For many years, the categorisation process has been unsatisfactory. The Prison Reform Trust runs an advice and information service for prisoners and their families, and categorisation queries and complaints typically rank amongst the most common that we receive. Delays, inconsistencies and allegations of unfairness all feature, as well as simple confusion about the rules that apply. An inefficient process also generates strategic problems for the management of the prison estate, where planning for the right type of prison places, in the right

quantity and in the right locations continues to present apparently insuperable difficulties.

5. The reform of the categorisation process that the prison service has undertaken is therefore welcome, and has been the subject of thorough and continuing consultation. The product of that work is a "Security Categorisation Policy Framework", updated in May 2021 and available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984877/security-categorisation-pf.pdf

6. The new policy includes the deployment of an advanced algorithm which draws on data from the whole of the prisoner's journey through the criminal justice system to produce a recommended initial categorisation, which can then be confirmed or overridden by the member of staff responsible for the process. This "digital categorisation service (DCS)" is described in the policy framework as follows:

The DCS is an online form and the mechanism for making, recording and justifying categorisation decisions. It supports staff in identifying relevant risk factors and reaching a decision on an individual's security needs but does not make categorisation decisions itself. It:

- *Provides a consistent framework for staff to record information about the various risk factors;*
- *Automatically highlights risk information to staff where available, to reduce the amount of manual checks required during the assessment process;*
- *Records previous categorisation assessments to support case management;*
- *At initial categorisation, provides a suggested categorisation as a starting point, which must then be reviewed by staff and changed where appropriate.*

7. There is much to welcome in the new policy framework and in the deployment of the algorithm. In particular, it holds out the prospect of a much more consistent application of the criteria that determine categorisation. In digitising the process, it also opens up the possibility of regular and more sophisticated monitoring, both locally and centrally. The new process was piloted and produced outcomes broadly similar to the previous system, suggesting that the algorithm is reflecting rather than radically changing expectations built up over many years.

8. However, the use of the algorithm does raise questions which we hope the committee might investigate.

9. The first, which will be a common theme, is that the algorithm depends on the quality of information fed into it. There is a risk that the undoubted benefits of consistency mask what is actually an entrenching of systematic unfairness. BAME and Muslim men are disproportionately represented in the prison population as a whole. But within prisons, that disproportion is even more marked in their representation in higher security categories. A careful and

admirably detailed Equalities Analysis carried out by the prison service for the categorisation policy framework, following a pilot phase, examined this phenomenon and found that the new process did not appear to be the cause of a small increase in the disproportionality of outcomes (once factors such as the location of the pilot prisons had been taken into account), but nor did it reduce it.

10. The unanswered question, therefore, is whether disproportionality is being driven by decisions and information from processes that precede the application of the categorisation algorithm and on which it depends. Specifically, intelligence about BAME and Muslim men which cannot be effectively challenged, and which may reflect conscious or unconscious bias may be driving unfair treatment which the categorisation algorithm then cements in a way which will follow the prisoner throughout their custodial history. Once information is in the system, regardless of its accuracy or continued relevance, it proves very hard to remove.

11. The second issue which the new process raises is the extent to which both Governors and the prison service corporately are equipped to interrogate the outcomes it produces, and in particular to both identify and investigate any indications that it is systematically unfair. The policy framework is alert to that possibility and gives Governors the following instruction:

"The Governor must nominate a manager whose responsibility it is to ensure that the categorisation/recategorisation process is functioning effectively; that decisions are fair, consistent and taken without bias; to provide quality assurance of decision making; to collect and analyse data in terms of protected characteristics (see paragraph 4.3) alongside other equalities data to ensure that there is a complete picture of any disproportionate impact, and to implement change where necessary."

12. But there is no indication of how local managers are to be equipped to carry out such an analysis, or what changes they might be empowered to make once they have.

13. In practice, for some prisons, the limited quantity of data available will make it hard to draw conclusions, certainly in the short term. The same cannot be said for national data, however, and whilst the equality analysis for the new policy framework is good in many respects, it leaves unanswered the question of why BAME and Muslim prisoners seem to fare less well both on initial categorisation and in the rarer cases where people are re-categorised to a higher security level.

14. To conclude, we suggest that this example from the prisons context highlights both the strengths and the weaknesses of using advanced algorithms in day to day decision making. The principal risk, in our view, is that the use of such algorithms locks into place systems which do not operate fairly. There is an inevitable tendency once such systems are operating to shy away from reforming them, even if the data they produce causes concern. To use a motoring analogy, managers and policy makers are likely to be less inclined to "look under the bonnet" when the technology they find there is unfamiliar. For the individuals who may have been treated unfairly, the chances of righting an individual wrong become even more remote.

15. We therefore suggest the following practical requirements for the future operation of this technology:

- An effective and easily accessed system for individuals to discover and be able to challenge the information which informs the algorithm in their particular case. Ultimately, there must be the possibility of a binding requirement from an external and independent arbiter to remove data from the system and re-assess a prisoner's category without it.
- Central advice for Governors on how to interpret their local data in order accurately to identify outcomes that are disproportionate and require explanation or reform.
- Regular national publication of the data the system produces, and a refusal to accept disproportionate outcomes which cannot be adequately explained.

16. The Committee may also wish to read an article published by the Police Foundation on this issue in 2020, which raises similar concerns, and is available at <https://www.police-foundation.org.uk/2020/10/greater-transparency-is-required-around-the-use-of-police-data-to-inform-decisions-in-prison/>

25 August 2021