Coin-flip judgement of psychopathic prisoners' risk

09 December 2013 by Seena Fazel

There are major flaws in the statistical tools for assessing whether people with serious psychological problems will reoffend, a psychiatrist says.

WOULD you be surprised to learn that many imprisoned people have serious personality disorders, such as psychopathy? Probably not. But I'd confidently predict you would be surprised to find that popular methods to predict the risk of reoffending to aid decisions on sentence length and release date for such people are no better than a coin toss.

This is the worrying verdict on long-established tools to assess the risk of violent reoffending, the latest blow to their utility.

Ironically, it was evidence that clinicians were failing to accurately predict the risk of violence in psychiatric patients that led researchers in the 1980s to use statistics to assess the chance of repeat offending, in a similar way to car insurers rating drivers. Checklists were devised that scored factors such as age at first arrest.

In the mid-1990s, Canadian researchers went beyond this actuarial approach and developed ways to identify risk that they argued were more tailored to offenders with mental disorders. These “structured clinical judgement” methods involve clinical judgement and a checklist.

Both approaches are increasingly used in high-income countries. This is fuelled, in part, by the need for defensible, transparent and consistent approaches to assessing risk, and by research that apparently demonstrates their superiority to subjective clinical decision-making. Many countries, such as the UK, use actuarial methods in probation service reports that influence sentencing decisions, and in the US, two-thirds of parole decisions include these methods.

Many US states use such tools to assess sexual offending risk and to help decide whether to exercise their powers to detain sexual offenders indefinitely after a prison term ends.

In England and Wales, these tools are part of the admission criteria for centres that treat people with dangerous and severe personality disorders. Outside North America, Europe and Australasia, similar approaches are increasingly popular, particularly in clinical settings, and there has been a steady growth of research from middle-income countries, such as China, documenting their use.

Until a few years ago, all looked good. By 2008, there were 40 systematic reviews and meta-analyses of the performance of these approaches, all of which found evidence that the assessments worked well. But many of the analyses were flawed. Problems included use of duplicate publications of the same study and inconsistent and clinically uninformative statistics. There was also an issue of bias, possibly unconscious, with the designers of assessment methods publishing papers on their accuracy that were consistently more glowing than those of independent groups.
So there was a need for new reviews, taking these issues into account. With colleagues, I have investigated the most commonly used instruments. We concluded in 2012 that they performed well for screening out low-risk offenders, or broadly identifying those who could benefit from additional interventions, but they were not accurate enough for making decisions about sentencing, release or preventative detention.

And a study published in September by Jeremy Coid and his colleagues at Queen Mary University of London, based on a group of released high-risk prisoners in England and Wales, found that the predictive accuracy of three widely used assessment tools was no better than chance for psychopathic offenders (British Journal of Psychiatry, DOI: 10.1192/bjp.bp.112.118471). Such people constitute about 10 per cent of the prison population, but repeat offend at a disproportionately high rate and more seriously. A possible explanation for the failing of the assessments for people with psychopathy is that they have inherently unstable, and unpredictable, lives. But it may also be that the tools, which perform only moderately for most offenders, will be poorer for all subgroups with different background risks.

In a sense, this new research has shown us where we stand. The assessment methods provide an expensive and moderately good aide-memoire, are prone to false positives, and are not accurate in providing individual probabilities of repeat offending. In practice, the high false-positive rate probably means that some offenders spend longer in prison and secure hospital than their true risk would suggest.

Of course, sensible clinicians and judges take into account factors other than the findings of these instruments, but their misuse does complicate the picture. Some have argued that the veneer of scientific respectability surrounding such methods may lead to over-reliance on their findings, and that their complexity is difficult for the courts. Beyond concerns about public protection, liberty and costs of extended detention, there are worries that associated training and administration may divert resources from treatment.

So what needs to be done? First, if there is a ceiling in the ability of these methods to predict serious crime, then new approaches are needed. Second, decision-makers must be told about the limitations. This is particularly important for those in the criminal justice system, where misuse of the tools is potentially more common due to the sheer number of offenders. Third, people assessing the accuracy of these tools must disclose any conflicts of interest, and research protocols must be registered to mitigate publication bias.

Violence risk assessment is an area where science and the law intersect, and thus it is especially important that it is built on firm foundations. The evidence supporting the field's tools needs independently funded research, careful evaluation by impartial experts, and uncompromising and repeated empirical scrutiny.

This article appeared in print under the headline "Coin-toss justice"

Seena Fazel is a Wellcome Trust senior research fellow at the department of psychiatry, University of Oxford, UK