editorial

ANTHONY MADEN

Standardised risk assessment: why all the fuss?

I have been surprised by the strength of feeling expressed by some opponents of standardised risk assessment. On the face of it, such opposition is a bizarre response to what amounts to nothing more than a special investigation. It is hard to imagine taking to the barricades in opposition to the Beck Depression Inventory, liver function tests or neuroimaging. The difference is that standardised risk assessment deals with violence and offending, so moral and emotional considerations intrude on scientific objectivity. In this editorial, I set out the principles underlying these investigations, deal with some of the objections to them and suggest ways in which they can be integrated into clinical practice.

Structured instruments

A structured risk assessment acts as an *aide-mémoire*, making sure that we collect all the relevant information. Many services have introduced a risk assessment form, partly as a talismanic charm against disaster, and the objective benefit of such forms is that they encourage staff to think about risk and to collect historical data.

While structured instruments help to avoid oversights, they have disadvantages. First, the forms irritate the people who have to fill them in. The temptation is to make the forms longer and longer for fear that something will be left out, which means that much of the material is irrelevant in any particular case, and the crucial facts might get lost in a mountain of dross. As well as the direct cost in staff time, there is also an opportunity cost because staff who are busy completing these forms are not doing other things, such as talking to the patient. It is rare to see any attempt being made to assess the impact of such forms.

The second problem is that structured instruments produce a large amount of information, but with no sense of what it all means. Best practice, therefore, is to use the data as the basis for a clinical team meeting. The team do the work of evaluation and planning, once the structured instrument has ensured that the necessary information is to hand. The outcome is, of course, only as good as the clinical team. If staff lack training or experience, they end up being overwhelmed by information that they are unable to use.

Standardised instruments

Most structured risk assessments are not standardised, but all standardised instruments are structured. Standardisation is carried out by giving a structured test to a population, to establish norms. This is where risk assessment gets interesting. The best analogy is with intelligence quotient (IQ) testing. It is moderately useful to know that one's patient is a bit slow in copying a geometric design, but the true power of IQ tests lies in ranking his or her performance alongside that of his or her peers. The same is true of risk. We need to know about a patient's criminal record, but the data are more powerful if his or her offending can be compared systematically with that of other patients.

Standardisation is simple in principle, but laborious and expensive in practice. The test must be applied to a large number of suitable people and the scores related to relevant outcome variables. Typically, these are offending in general, or sexual and violent offending in particular. Other outcome variables include failures of conditional release or non-compliance with out-patient treatment. Standardised scores allow us to position a patient on a scale. Does he or she resemble the easier or the more difficult end of the prison or secure unit population? Does he or she score like those patients who went on to re-offend after discharge or like those who led stable and settled lives? This approach is atheoretical and identical to the process used by insurance companies in assessing risk. Hence, standardised instruments are also known as actuarial measures.

In addition to large numbers, standardisation requires a population chosen to minimise the possibility of bias, arising from age, gender, ethnic origin or some other variable. One of the obstacles to the adoption of such instruments in the UK is that most of the standardisation has been done in Canada and the USA and we cannot assume that the same norms will apply here, although initial results using the revised Hare Psychopathy Checklist (PCL–R) (Hare, 1991) in UK prison populations are encouraging. The rank ordering of individuals is similar, even if the absolute scores are not the same.

Compared with the process of standardisation, the content of the risk assessment instruments may be of less importance. Actuarial instruments rely heavily on previous offending, with the differences between

201

instruments being seen in the way the information is coded or in the inclusion of specialised material, such as that related to sex offending or mental illness. For example, the Violence Risk Appraisal Guide (VRAG) (Harris *et al*, 1993) and the Historical Clinical Risk 20 (HCR–20) (Webster *et al*, 1995) both include a measure of psychopathy derived from the revised Hare Psychopathy Checklist. The Sex Offender Risk Appraisal Guide (SORAG) is a modification of the Violence Risk Appraisal Guide, designed to address specific sexual risk.

Risk scales have proliferated in recent years because it is easy to create new ones by adding items to the core offending history. This process appeals to the vanity of academics but it destroys the benefits of standardisation, which must begin all over again with the new scale. Rather than search for the Holy Grail of the perfect risk assessment instrument, there is a strong case for accepting the flaws of an existing scale, which are often outweighed by the benefits of standardisation.

Problems with standardised risk assessment

British psychiatrists remain suspicious of these scales for a variety of reasons, some more rational than others. I will now discuss some of the real and alleged problems surrounding their use.

Fear of unemployment

The Luddite position, though rarely made explicit, is that clinical skills will become obsolete if use of these scales is allowed to spread. This objection is unacceptable in principle if the aim is to improve risk management rather than to provide sheltered employment for clinicians. It is also untenable in practice. While some enthusiasts have argued that actuarial methods should supplant clinical estimation of risk (Quinsey *et al*, 1998, p. 171), a more balanced review (Monahan *et al*, 2001, pp. 129–136) concludes that the proper place of such instruments is an adjunct to good clinical practice.

Stigmatisation

There is a valid fear that, because of the mystique attached to some of these tests, patients given a high score will be rejected by services or held in detention for unnecessarily long periods. The danger is greatest in relation to psychopathy, where a high score has been used to exclude offenders from treatment programmes on the grounds that treatment might make them worse. The evidence on this point is far from conclusive, and the studies showing no benefits from institutional treatment also show that supervision reduces re-offending rates in high scorers, in much the same way as with other offenders.

The stigmatisation issue relates to the novelty of the tests. Again, the analogy with IQ testing is useful. In the early days, clinicians gave too much weight to intelligence tests, compared with other aspects of a case. This led to a backlash in which psychologists would refuse to report

IQ scores for fear of the damage that might be done to the patient's treatment. The area has now stabilised, with recognition of both the benefits and limitations of IQ testing. It is reasonable to hope that the same sense of perspective will emerge in standardised risk assessment, once the dust settles.

In any case, the stigmatisation argument verges on the frivolous when dealing with a high risk of violent or sexual offending. Such offenders are stigmatised by the risk that they pose to other people, not by a scale that makes that risk explicit. The correct response is to develop better methods of management rather than to adopt the irresponsible, 'ostrich-like' tactic of refusing to even measure or acknowledge the risks.

A better choice of prejudices?

Opponents of standardised risk assessment tend to assume that the test score will be high and will allocate a patient to a risk category unwarranted by clinical data. However, it is important to consider the opposite scenario. Psychiatry has a bad record of detaining patients in excessive security. All those patients who are held inappropriately in high-security were put there by doctors exercising unfettered clinical judgement. Such patients deserve a proper, standardised assessment of risk. The score should not lead automatically to any action, but provide an objective basis for debate, far superior to the consultant's rule of thumb.

Similarly, forensic psychiatry has to take seriously the statistical over-representation of patients from ethnic minorities in all locked settings and the over-representation of women in high-security. Most of these patients were locked up by White male doctors and any objective evidence of risk should therefore be welcomed. I am not persuaded that there is widespread racism or sexism in forensic psychiatry, but we cannot expect people simply to take our word for it.

Mental illness *v.* personality disorder

In general, prediction from actuarial scales is easier in the case of personality disorder than in mental illness. In the MacArthur study of 1000 general psychiatry patients, psychopathy emerged as the best single predictor of violence (Monahan *et al*, 2001, p. 65–72). This is probably because of the relative stability of the traits that make up antisocial personality disorder or psychopathy.

Once psychosis supervenes, most bets, if not all of them, are off. Standardised assessments will identify the higher risk associated with comorbid personality disorder or substance misuse, but they are disappointing when used to predict violence in uncomplicated psychosis.

The absence of psychopathy is no guarantee that a person with psychosis will not offend as a direct consequence of that psychosis. Anecdotal evidence suggests that there is an inherent unpredictability about psychotic violence that is not associated with comorbidity. We also know that a considerable proportion of new cases of schizophrenia present with violence. There is no scope for

prediction when the target event occurs before the diagnosis has become apparent.

Static *v.* dynamic predictors

Following on from the discussion about mental illness *v.* personality disorder, there is a wider point concerning the relative value of static and dynamic variables as predictors of violence. Static factors include offending history, juvenile delinquency and previous treatment failure or non-compliance. Dynamic factors include substance misuse, mental state and mixing with criminal associates. The best scales rely heavily on static factors. This means that they can get the patient into trouble (or hospital) but they cannot get him or her out, in that treatment will never change an individual's past. In this respect, these instruments fail to meet the needs of the clinician whose main interest is in knowing when patients are ready for discharge.

This is a problem, but surely not a reason for abandoning such assessments. We need to know the bad news so that we can target resources. We need also to spend more time assessing the value of dynamic variables. It may well be that static factors have assumed such importance merely because they are easier to measure and study. The priority for future research must be to include dynamic variables in risk assessment instruments. An example of this approach is found in the Violence Risk Scale (Wong & Gordon, 2000), which has six static and 20 dynamic variables.

Limits to prediction

Many critics have impossible expectations, the most unrealistic being that a scale can indicate what a patient is going to do tomorrow, next week or next year. When these optimists find that the crystal ball does not work, they want to throw it in the bin. The fault lies with the clairvoyant fantasy. At best, these tests can only assign patients to a broad category that is defined by a statement of probability.

For example, it may be possible to state that a patient has characteristics implying a 10% chance of re-offending during the next 5 years. When such information is used in deciding whether or not to detain a patient, the point at issue is whether that risk is acceptable. There is no implied knowledge of what the patient will or will not do in the future. Also, the score does not tell us what to do – that decision rests on a complex balancing of a patient's interests against the safety of others. Critics tend to assume that detention is the only response to high risk, whereas a range of community options is effective in reducing risk.

Doctors have little experience of working explicitly with probability and they are not very good at it. The terminology of signal detection theory has been misused to argue that a 10% risk involves detaining nine false positives for every true one, resulting in the test having no value. But these instruments do not claim to identify offenders in advance, only to make statements of probability. It is reasonable and ethical for society to conclude

that a given risk of offending is unacceptable and to detain the individual because of that risk. A discussion about the costs of this approach to crime prevention should follow, but the moral principle is uncontroversial. We have laws against speeding, not because of any certainty that a particular driver will have an accident, but because the probability of accidents is unacceptably high. A similar principle applies to violence associated with mental disorder.

Low-risk populations

Once the level of risk in a population falls to a low level, as in many general psychiatric populations, actuarial instruments are less useful. In such a population, most violence will be by patients with low scores. After all, 100 patients with a 1% risk add up to one event, as do two patients with a 50% risk. In this setting, actuarial instruments may be useful in screening out a high-risk minority, but they will not provide useful data on most patients.

A compromise – the place of standardised risk assessment

Standardised or actuarial risk assessment is not an alternative to clinical skills, but it should be used to improve clinical practice. The extent to which it will be useful depends on the context.

In general psychiatry, most services will want a minimum data set, amounting to a simple, structured assessment to inform care planning. Although there is no consensus, one would hope that such an assessment would pick up comorbid substance abuse and personality disorder, which are the main factors increasing the risk of violence in such populations. The case for further, standardised assessment is weak and must be set against the direct and opportunity costs. In a service dealing mainly with mental illness, resources may be better expended on improving compliance in the whole population rather than attempting to target high-risk individuals.

A history of actual or threatened violence shifts the balance in favour of further, standardised assessment. The nature of that assessment will depend on the extent of the apparent risk and should probably be decided in conjunction with forensic services, either on a case-by-case basis, or as a matter of policy for liaison between the two services.

The cost–benefit equation is different within forensic services, where a battery of standardised assessments should be routine, certainly for in-patient services. As cost per day and length of stay are both high, so the relative cost of the assessment falls to the point of insignificance. Also, as forensic services are relatively privileged in terms of resources, it is reasonable that they should justify their extra resources by showing that they take patients who present greater risks. It is not possible to do sensible research on outcome without an objective measure of the baseline risk.

Finally, in specialised areas such as the treatment of sex offenders and those with severe personality disorder

editorial

editorial

or psychopathy, standardised assessments should be mandatory in all cases.

Declaration of interest

None.

References

HARE, R. (1991) *The Hare Psychopathy Checklist — Revised.* Toronto: Multi-Health Systems.

HARRIS, G., RICE, M. & QUINSEY, V. (1993) Violent recidivism of mentally disordered offenders. The development of a statistical prediction instrument. *Criminal Justice and Behaviour*, **20**, 315.

MONAHAN, J., STEADMAN, H. J., SILVER, E., *et al* (2001) *Rethinking Risk Assessment. The MacArthur Study of Mental Disorder and Violence.* Oxford: Oxford University Press.

QUINSEY, V., HARRIS, G., RICE, M., *et al* (1998) *Violent Offenders:* *Appraising and Managing Risk.* Washington, DC: American Psychological Association.

RICE, M. E. & HARRIS, G. T. (1997) Cross validation and extension of the *Violence Risk Appraisal Guide* for child molesters and rapists. *Law and Human Behavior*, **21**, 231–241.

WEBSTER, C., DOUGLAS, K., EAVES, D., *et al* (1995) *HCR−20 Assessing Risk for Violence (Version 2).* Vancouver: Simon Fraser University.

WONG, S. & GORDON, A. (2000) *Violence Risk Scale.* Saskatoon: Regional Psychiatric Centre.

Anthony Maden Professor of Forensic Psychiatry, Imperial College London, Academic Centre, Trust Headquarters, WLMH NHS Trust, Uxbridge Road, Southall, Middlesex UB1 3EU

204