

## Correspondence

Edited by Kiriakos Xenitidis and  
Colin Campbell

## Contents

- Shall evidence-based risk assessment be abandoned?

### Shall evidence-based risk assessment be abandoned?

Hart *et al*<sup>1</sup> argued that actuarial risk assessments (Violence Risk Appraisal Guide, VRAG, and Static-99) are 'virtually meaningless'. They committed statistical error by misapplying confidence intervals (see <http://www.mhcp-research.com/hmcrespond.htm> or <http://bjp.rcpsych.org/cgi/eletters/190/49/s60>). Confidence intervals do not capture the 'precision' of individual scores. The appropriate statistic is standard error of measurement, which for the VRAG indicates that any individual score carries less than 0.05 probability of misclassification by more than one category.<sup>2</sup>

Hart *et al* also erred in using 'precision' and 'accuracy' as synonyms. Accuracy is best assessed by sensitivity, specificity and their trade off. More than 40 evaluations of the VRAG (and the allied Sex Offender Risk Appraisal Guide) in approximately 8000 released prisoners, sex offenders, forensic patients, compulsorily admitted ('civil') psychiatric patients and other samples from seven countries have employed follow-ups from several weeks to 10 years. Predictive effects (in sensitivity–specificity terms) are large. Contrary to Hart *et al*, scores predict the speed and severity of recidivism. Most commonly, there are no statistically significant differences between observed rates and expectations based on norms,<sup>3</sup> especially given variation predicted by Bayes' Rule. Hart *et al*'s statistical argument does not and cannot refute empirical results supporting the accuracy of actuarial risk assessments in predicting the violent recidivism of offenders.

The notion that it is scientifically wrong to base individual decisions on such post-analytic groupings (inappropriately referred to as 'group data' by Hart *et al*) has been thoroughly refuted.<sup>2,4</sup> What should a clinician do when considering the release of one previously violent forensic patient? Hart *et al* implied that no decision should be made, recommending concentrating on subjectivist issues instead. We disagree. An actuarial tool (such as the VRAG or Static-99) is simply an efficient distillation of relevant empirical evidence. Actuarials do not afford certainty, but, as Hart *et al* acknowledged, are more valid than any other method. The undeniable superiority of actuarials means that their use can optimise the balance between public safety and offenders' civil liberties. Hart *et al*'s advice to eschew risk-related decisions means less accurate decisions that cumulate in avoidable harm to victims, unnecessary restriction of offenders, or both.

1 Hart SD, Michie C, Cooke DJ. Precision of actuarial risk assessment instruments. *Br J Psychiatry* 2007; **190** (suppl 49): s60–5.

2 Quinsey VL, Harris GT, Rice ME, Cormier CA. *Violent Offenders: Appraising and Managing Risk* (2nd edn). American Psychological Association, 2006.

3 Harris GT, Rice ME. Characterising the value of actuarial violence risk assessment. *Crim Justice Behav* 2007; in press.

4 Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychol Public Policy Law* 1996; **2**: 293–323.

**Grant T. Harris**, Director of Research, Mental Health Center Penetanguishene, 500 Church Street, Penetanguishene, ON L9M 1G3, Canada. Email: gharris@mhcp.on.ca; **Marnie E. Rice**, Mental Health Center Penetanguishene, Penetanguishene, Canada; **Vernon L. Quinsey**, Department of Psychology, Queen's University, Kingston, Ontario, Canada

doi: 10.1192/bjp.192.2.154

**Author's reply:** Harris *et al* make a number of claims:

- That we misapplied confidence intervals to actuarial test scores: in fact, we used confidence intervals to evaluate the estimated probability of violence associated with test scores, not the raw scores themselves. The (many) problems with raw scores on actuarial tests are a separate issue.
- That we used 'precision' and 'accuracy' synonymously. We did not – we simply recognised the important association between these concepts: the accuracy with which actuarial tests can predict future violence in an individual case depends on the precision of group data. As every research trainee learns, reliability places an upper bound on validity.
- That their views about basing individual decisions on group data are supported by Grove & Meehl.<sup>1</sup> They ought to read Grove & Meehl more carefully:

'There is a real problem, not a fallacious objection, about uniqueness versus aggregates in defining what statisticians call the reference class for computing a particular probability in coming to a decision about an individual case'<sup>1</sup> (p. 306)

Grove & Meehl's (lengthy) discussion, which includes the issue of the precision of group estimates, is echoed in our paper.

- That their belief in the 'undeniable superiority of actuarials' is supported by Grove & Meehl.<sup>1</sup> However, Harris *et al* continue to confuse group and individual data. Grove & Meehl concluded that actuarial decision-making was superior to clinical judgement in about 45% of the studies that they reviewed; in the others, clinical judgement was equally accurate or even more accurate. Put differently, the 'on average' superiority of actuarials translated into superiority in slightly less than half of the individual comparisons. This is an important trend, obviously, but hardly a sound basis for high-stakes gambling on one outcome. As good scientists, we recommend against betting big on the toss of a single coin.

We strongly support evidence-based practice, but Harris *et al* have confused 'evidence-based' with 'statistically based'. They should recognise that in forensic mental health, as in many areas of life, good practice does not equate to mindless reliance on simplistic statistical algorithms.

1 Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychol Public Policy Law* 1996; **2**: 293–323.

**Stephen D. Hart**, Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6. Email: hart@sfu.ca; **Christine Michie**, **David J. Cooke**, Glasgow Centre for the Study of Violence, Glasgow Caledonian University, Glasgow, G4 0BA, UK

doi: 10.1192/bjp.192.2.154a