Shall evidence-based risk assessment be abandoned?

Hart et al. 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.

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, 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. They ought to read Grove & Meehl’s (lengthy) discussion, which includes the issue of the precision of group estimates, is echoed in our paper.

(a) 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.

(b) 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.

(c) That their views about basing individual decisions on group data are supported by Grove & Meehl. 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’ (p. 356)

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

(d) That their belief in the ‘undeniable superiority of actuarials’ is supported by Grove & Meehl. 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

2 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.


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.

Shall evidence-based risk assessment be abandoned?
Grant T. Harris, Marnie E. Rice and Vernon L. Quinsey
Access the most recent version at DOI: 10.1192/bjp.192.2.154

References
This article cites 2 articles, 1 of which you can access for free at:
http://bjp.rcpsych.org/content/192/2/154.1#BIBL

Reprints/permissions
To obtain reprints or permission to reproduce material from this paper, please write to permissions@rcpsych.ac.uk

You can respond to this article at
/letters/submit/bjprcpsych;192/2/154

Downloaded from
http://bjp.rcpsych.org/ on April 14, 2017
Published by The Royal College of Psychiatrists