dehydration and increased cardiovascular output – particularly important in the elderly or those with pre-existing disease. A range of antipsychotic drugs are known to inhibit sweating and therefore thermoregulation. Recent work has shown that deaths from respiratory and external causes are particularly increased at high temperatures (Hajat *et al*, 2007). Further research is needed on the pathophysiology of heat, but it is clear that persons with mental illness remain a high-risk group for heatwave mortality (Kovats & Ebi, 2006).

Bouchama, A. & Knochel, J. (2002) Heat stroke. *New England Journal of Medicine*, **346**, 1978–1988.

Hajat, S., Kovats, R. S. & Lachowycz, K. (2007) Heat-related and cold-related deaths in England and Wales: who is at risk? *Occupational and Environmental Medicine*, **64**, 93–100.

Kovats, R. S. & Ebi, K. L. (2006) Heatwaves and public health in Europe. *European Journal of Public Health*, **16**, 592–599.

L. A. Page Department of Psychological Medicine, Institute of Psychiatry, King's College London, Weston Education Centre, London SE5 8RJ, UK. Email: l.page@iop.kcl.ac.uk

S. Hajat, R. S. Kovats Public and Environmental Health Research Unit, London School of Hygiene and Tropical Medicine, London, UK

doi: 10.1192/bjp.191.6.560a

Avoiding errors about 'margins of error'

When discussing actuarial risk assessment instruments (ARAIs), Hart *et al* (2007) acknowledge that 'prediction' may refer to probabilistic statements (e.g. a 'prediction' that an individual 'falls in a category for which the estimated risk of violence was 52%': p. s60). For unclear reasons, however, the authors seem to value only predictions with right or wrong outcomes. They therefore regard statements about future behaviour of large groups (where one can be almost certain that the fraction of persons who act a certain way will fall within a narrow range of proportions) as potentially 'credible', but predictions for individuals as meaningless.

If the purpose of risk assessment is to make choices, then well-grounded probabilistic predictions about single events help us. Suppose we conclude that it is legally and ethically acceptable to impose preventive confinement upon individuals in ARAI categories with estimated recidivism rates above a specified threshold. This policy entails making 'false-negative' and 'false-positive' decision errors. We recognise, however, that unless we are omniscient perfection is not an option and ARAIs simply help us make better decisions than we otherwise could.

How do 'margins of error' in estimated recidivism rates affect our decision process? Hart *et al* believe their 'group risk' and 'individual risk' 95% confidence intervals speak to this problem. Their group intervals are standard confidence intervals for estimated population proportions based on random samples. If the threshold lies outside the group risk confidence interval for a category, then we can be reasonably certain that a decision we make concerning someone in that category is the same decision we would make if we knew the true recidivism rate for that category. If the threshold falls within a category's group risk confidence interval, then our estimate quite possibly might lead to the 'wrong' decision. Statistical decision theory (Berger, 1985) shows, however, that it is still a sensible strategy to choose whether to confine a member of a category based on which side of the threshold our estimated risk falls.

Hart *et al* talk about 'individual risk' as though it is something different from category (or 'group') risk. Yet if all one knows about an individual is his membership of a risk group, what can 'individual risk' mean? The authors do not say. If 'individual risk' refers to believed-to-exist-but-unspecified differences between individuals within a category, such differences should not affect choices by a rational decision-maker. The 95% CIs for 'individual risk' pile nonsense on top of meaninglessness. Hart *et al* describe the replacement of '*n*' by '1' in the Wilson (1927) formulae as 'ad hoc', but this substitution makes no sense when the basis for the estimated proportion is an *n*-member sample. With '1' in place of '*n*', the formulae just don't mean anything.

Using ARAIs raises serious moral problems as well as the valid scientific questions that Hart *et al* mention. But in faulting the capacity of ARAIs to address an unspecified quantity called 'individual risk', and in dressing up this notion with misapplied formulae for confidence intervals, Hart *et al* ultimately create a muddle.

Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis* (2nd edn). Springer-Verlag.

Hart, S. D., Michie, C. & Cooke, D. J. (2007) Precision of actuarial risk assessment instruments. Evaluating the 'margins of error' of group *v.* individual predictions of violence. *British Journal of Psychiatry*, **190** (suppl. 49), s60–s65.

Wilson, E. B. (1927) Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**, 209–212.

D. Mossman Department of Psychiatry, Wright State University Boonshoft School of Medicine, 627 S. Edwin C. Moses Boulevard, Dayton, Ohio 45408-1461, USA. Email: douglas.mossman@wright.edu

T. Sellke Department of Statistics, Purdue University, West Lafayette, Indiana, USA

doi: 10.1192/bjp.191.6.561

Authors' reply: Actuarial risk assessment instruments (ARAIs), constructed using data from known groups, are used to make life-and-death decisions about individuals. How precisely do they estimate risk in individual cases? The 95% CI for proportions, which evaluates the precision of risk estimates for ARAI groups, cannot be used for individual risk estimates unless one makes a very strong assumption of heterogeneity – that ARAIs carve nature at its joints, separating people with perfect accuracy into non-overlapping categories. No one, not even those who construct ARAIs, makes this assumption. So, we ask again, what is the precision of individual risk estimates made using ARAIs?

Mossman & Sellke criticise us for inadequately defining 'individual risk' and for using an ad hoc procedure to estimate the margin of error for individual risk estimates, which they opine served only to 'pile nonsense on top of meaninglessness'.

We must plead guilty to some of the charges levelled by Mossman & Sellke – indeed, we did so in our paper, acknowledging the conceptual and statistical problems with the approach we used. In our defence, we claimed duress: because developers used inappropriate statistical methods to construct ARAIs, we could not use appropriate methods to evaluate them. Violent recidivism was measured in the ARAI development samples as a dichotomous, time-dependent outcome, and so the developers ought to have used logistic regression or survival analysis to build models; if they had, one could directly calculate logistic regression or survival scores for individuals and their associated 95% CIs.

But we also plead that these charges are irrelevant to our conclusion. As we discussed, to reject our findings that the