MASTERS, C. L. & RICHARDSON, E. P. (1978) Subacute spongiform encephalopathy (Creutzfeldt-Jakob disease): the nature and progression of spongiform change. *Brain*, **101**, 333–334.

PERRY, R. H., IRVING, D., BLESSED, G., *et al* (1990) Senile dementia of Lewy body type. A clinically and neuropathologically distinct form of Lewy body dementia in the elderly. *Journal of Neurological Sciences*, **95**, 119–139.

SMITH, T. W., ANWER, U., DEGIROLAMI, V. H., *et al* (1987) Vacuolar change in Alzheimer's disease. *Archives of Neurology*, **44**, 1225–1228.

E. JANE BYRNE
J. LOWE

University of Nottingham Medical School
Queen's Medical Centre
Nottingham NG7 2UH

'Beam them up, Scotty!'

SIR: Harrison and Roberts are to be congratulated on drawing on research data from the future as well as the present and past (*Journal*, April 1991, **158**, 457–470). If we are to proceed to this future, however, its integrity must be maintained by not tampering with it in the present. The finding 'life, Jim, but not as we know it' was reported by Science Officer Spock and not Chief Medical Officer McCoy. I would hesitate to suggest it was a Freudian slip on the part of a medic who made this mistake, but instead would remind us of Mr Spock's words of hope for psychiatry when visiting the Galaxy's remaining asylum for the criminally insane on Elba II (Star Trek, Episode 71, *Whom Gods Destroy*): "A total of 15 criminally insane out of billions is not what I would call an excessive number".

JACQUELINE M. ATKINSON

University of Glasgow
Department of Public Health
2 Lilybank Gardens
Glasgow G12 8RZ

The Yates' correction factor in chi-squared analyses

SIR: On page 236 of Healy *et al*'s paper (*Journal*, February 1991, **158**, 234–237), it says, under the heading "Data analysis", that data collected on the different groups was compared using the χ^2 test with Yates' correction factor.

The correction factor originally devised by Yates (1934) was applied to Pearson's chi-squared to improve the approximation to a continuous function. Standard statistical texts used by students over the years have given this correction factor as a matter of course (e.g. Blalock, 1979). However, this correction factor has been questioned in recent years and Hopkins & Glass (1978) discuss this. Research findings show that the chi-squared works well even when the average expected frequency is as low as 2. Hopkins & Glass quote Camilli & Hopkins (1978) to say that not only is the Yates' correction for continuity unnecessary, it also causes the already conservative values for chi-squared to be even more conservative.

It is true that statisticians do not all agree on these points; Yates himself restated his arguments as recently as 1984. If, however, we are concerned with the degree of conservativeness of the values obtained, research workers might do well to analyse data without the Yates' correction factor, or use both methods.

BLALOCK, H. M. (1979) *Social Statistics, 2nd edn*. Maidenhead, UK: McGraw-Hill Book Company International Student Edition.

CAMILLI, G. & HOPKINS, K. D. (1978) Applicability of chi-square to 2 × 2 contingency tables with small expected frequencies. *Psychological Bulletin*, **85**, 163–167.

HOPKINS, K. D. & GLASS, G. V. (1978) *Basic Statistics for the Behavioural Sciences*. Englewood Cliffs: Prentice-Hall.

YATES, F. (1934) Contingency tables involving small numbers and the Chi-square test. *Journal of the Royal Statistical Society* (suppl. 1), 217–235.

YATES, F. (1984) Tests of significance for 2 × 2 tables. *Journal of the Royal Statistical Society*, Series A, **147**.

PETER DAWSON

Mental Health Services
Whakatane, Bay of Plenty
New Zealand

The strength of association

SIR: May I briefly underline an important point made by Muijen (*Journal*, May 1991, **158**, 713). He rightly stresses that it is the size of a correlation (provided it is significant) which measures the strength of the association between the two variables. But how large must a correlation be, to be considered useful? An answer to this thoroughly practical question can be derived from a theorem of information theory (Pinsker, 1964, p. 123), which deserves to be better known.

The usefulness of a correlation lies essentially in its predictive power. If two variables are correlated and we know the value of one of them on a given occasion, we know something about the value of the other. The higher the correlation, the more we know, i.e. the more information it provides. The amount of information is given by:

$$I(x,y) = -\tfrac{1}{2}\log(1-r^2)$$

It is expressed in binary units ('bits') if logarithms to base 2 are used.