LETTER TO THE EDITOR

SF-36 health status questionnaire

Sir – You published a paper by Sue Ziebland on the usefulness of the SF-36 health status questionnaire for measuring health gain in population surveys.1 She concluded that the "variation in responses in a general population make it an inadequate tool for assessing the diffuse impact of health interventions directed at the whole community".

We disagree with this. We do not underestimate the difficulties in using SF-36 for general population monitoring; but we do not believe that this potential use can be rejected on the grounds given by Ziebland. At Midland Health in New Zealand, we have been testing the SF-36 on our population, and found illuminating the information which it provides on age, gender, and ethnic (Maori-Maori) differences in health status.2 Naturally, we are keenly interested in its possible use for monitoring changes in these differences over time.

Ziebland's argument is that changes in a population's self reported health status should not be assessed in terms of the statistical significance of the change, but rather in terms of the "effect size". (The effect size being simply the change in mean score divided by the population standard deviation (SD) of individual scores. An effect size of 0.20 is labelled as "small"; one of 0.40 as "moderate"; and one of 0.80 as "large").

Ziebland calculates (using data from the Oxford health and lifestyle survey) the effect sizes of hypothetical improvements in health status. For example, if mean scores of men and women in social class IV were increased to match those in social class I. The effect sizes she gets are generally "small" or "negligible". Hence her conclusion that "even if policies did improve the health status of those in social class V to that of social class I, the SF-36 would not be equal to the task of measuring the outcome". This is for a increase in the average score on "general health" from 70.3 to 75.1. Given a population standard deviation of 21.2, the effect size is labelled as "small".

If, however, standard statistical tests are applied, this result would be regarded as highly significant (the sample size is 236).

One reason for Ziebland's results is that her initial population - class V - is one of particularly high variability - an SD of 21.2 compared with 17.8 for the class I population. Another reason is that variability in the data is increased by the presence of other determinants than social class. In particular the effects of age would account for a good part of the variability in individual scores in the social class datasets. In practice the preferred approach would, in any case, be to try to measure change in relatively homogeneous population subgroups.

Our fundamental objection in Ziebland's paper is, however, to her proposition that "effect size" is the proper measure for monitoring the significance of a population change in health status, rather than the statistical significance of the change. If accepts 70.3 would seem to rule out the use of SF-36 for measuring significant, but small to moderate, changes in average health status in many populations. That would be both unfortunate, and unnecessary. The standard tests of statistical significance still retain their relevance for measuring change in population health status, no matter how variable the population from which a given sample is drawn.

2 Kokaia J, Wheaon M, Seas J. Validation of the SF-36 health survey on a New Zealand population. Midland Regional Health Authority, PO Box 1031, Hamilton, New Zealand.

NOTICES


Vth Annual British Epidemiology Society (BEES) Workshop, 12 January 1996 at the University of Nottingham. For further information, contact: Melanie Bowesman, Secretary to Dr Hywel Williams, Department of Dermatology, C Floor, South Block, Queen's Medical Centre, Nottingham, NG7 2UH. Tel: 0115 942 1421 ext 44539. Fax: 0115 970 9003.