the long-term efficacy was poor and the CGI severity results were even in favour of placebo at some later weeks.

One other problem not discussed by the authors might be the absence of a correction for centre effects. There were 37 centres in 6 different countries participating in this trial, varying from 1 centre in Finland to 13 centres in France. The question concerning centre effects, which might be anticipated in such a multicentre trial, was not discussed. Apparently nothing about minimum or maximum number of enrolments per centre was written in the protocol. From our own experience we can state that multinational studies in psychiatry in Europe are not easy to organise and conduct. Furthermore, it would be very helpful to know how many international and national training sessions have been organised as well as data concerning the inter-rater variability.

Data from studies without the definition of the major outcome variables *a priori*, should not be accepted as final proof of efficacy. Therefore we tend to see the study of Doogan & Caillard merely as a feasibility and hypothesis-generating study.

COMMISSION OF THE EUROPEAN COMMUNITIES (1989) *The Rules Governing Medicinal Products in the European Community*, III, 209–218.

LEE, J. H. (1990*a*) Clinical review of efficacy data. In *NDA-19-839 Sertraline: Safety and Efficacy Considerations for Use in the Management of Depression*, pp. 7–19. Psychopharmacological drugs advisory committee, Rockville, USA: Food and Drug Administration.

— (1990*b*) *Summary Basis of Approval: Statistical Review and Evaluation of Sertraline HCL; NDA #19–836*. Rockville, USA: Food and Drug Administration.

JAN M. KEPPEL HESSELINK
PERRY M. C. DE JONGH

Medical Department
Bayer Netherlands
Nijverheidsweg 26, Mijdrecht
The Netherlands

AUTHOR'S REPLY: Our study was reviewed by the Food and Drug Administration, and a number of methodological matters were discussed.

We accept that there was no absolute *a priori* definition of responder mentioned in the protocol. However, all the usual criteria for response were applied in the analysis of this study. Irrespective of which criteria were used, the result always significantly favoured sertraline over placebo. Thus it is not appropriate to suggest that the data analyses were designed arbitrarily.

A key criticism was that the excess rate of discontinuation of placebo patients over sertraline did not allow the use of an observed-cases analysis to adequately assess drug effect. In a maintenance study, patients remaining well will continue in the study. Therefore, comparisons of CGI severity between sertraline and placebo are unlikely to show any significant difference. The most meaningful statistical analysis is the Kaplan-Meyer survival estimate, which is a conventional analysis used in such situations. This analysis, which controls simultaneously for drop-outs, shows superiority of sertraline over placebo at all time points. It is our firm belief that observed-cases analyses are inappropriate at these time points.

One item not discussed in the paper was the analysis of centre effects. This was investigated and no significant treatment by centre interaction was identified. Thus the number of centres was not a significant factor affecting results. Further, we believed it was unnecessary to conduct inter-rater reliability sessions when the key efficacy measure was Clinical Global Impression. Inter-rater reliability is more to be considered when discrete rating scales, such as Hamilton or Montgomery-Åsberg scales, are being used.

This study was an ambitious project to identify if there was any benefit in maintaining patients long term on sertraline treatment. The conclusions of this study remain that sertraline is of benefit in the long term for controlling relapse of depression.

D. P. DOOGAN

Pfizer Ltd
Sandwich
Kent CT13 9NJ

SIR: I have had the opportunity of independently reviewing the data from the sertraline placebo long-term treatment study and my conclusions have been published (Montgomery *et al*, 1991). The striking finding in the study was that it did not matter which relapse criteria were adopted since there was a significant advantage for sertraline over placebo with the measures that I examined using either the Hamilton Depression scores or the Clinical Global Severity scale.

The criticism that the analysis was made on *post hoc* definitions of relapse is valid as was discussed in our paper. There is debate as to which relapse criteria are most sensitive to long-term treatment effect. The sertraline–placebo database provides one of the few chances of comparing the effect of different relapse criteria.

The efficacy of an antidepressant in long-term treatment is measured by its ability to reduce the number of relapses or recurrences compared with placebo. The long-term treatment studies do appear

to be a more robust way of testing efficacy and one not as vulnerable as the acute placebo-controlled studies to centre effects, or variable placebo response rates. It is for this reason that many of us turn to long-term treatment studies for proper reassurance about the efficacy of an antidepressant. I have no doubts from the results of the study that sertraline is indeed effective in long-term treatment.

MONTGOMERY, S. A., DOOGAN D. P. & BURNSIDE R. (1991) The influence of different relapse criteria on the assessment of long term efficacy of sertraline. *International Clinical Psychopharmacology*, **6** (suppl. 2), 37–46.

STUART MONTGOMERY
Academic Department of Psychiatry
St Mary's Hospital
Praed St
London W2 1NY

Reports of the death of factor analysis are greatly exaggerated

SIR: Bech *et al* (*Journal*, February 1992, **160,** 206–211) presented psychometric analyses of the Hamilton scales for depression and for anxiety, and the SCL–90. The results, from a large and rigorous study, are fascinating and have major implications for cross-cultural psychiatric research. However, a subplot in the paper appears to be a "head-to-head" comparison of the merits of two psychometric methods: item-response analysis using the Rasch model versus exploratory factor analysis. This subplot is revealed in lines in the discussion: "Factor analysis is still considered to be an important psychometric method, but this study has demonstrated the difficulties in interpreting the results of such analysis." (pp. 209–210) and "The advantage of latent structure analysis has been demonstrated in this study." (p. 210). This last sentence continues a widely used misnomer as both Rasch model-item analysis and factor-analytic methods seek to reveal latent structure in item data: the former can reveal non-linear item-response parameters on a single latent dimension of variation (e.g. depression, anxiety, general psychological distress); the latter can reveal multiple linear dimensions in data at the cost of assumptions about linearity of response that are not made in Rasch model-item analysis.

The presentation in the paper oversimplifies the situation: item analyses can use non-linear item responses to reveal the fit between the data and a model which assumes a single underlying latent dimension. Furthermore, these methods can test the homogeneity of the scaling in relation to apparent position on that dimension or in relation to other variables such as the counties in which the study took place. As all multiple-choice responses must be linear these are desirable abilities. However, current Rasch methods cannot reveal multidimensional latent variable structure. This is relevant in psychiatric scales such as the HRSD where there might be dimensions not only of severity of depression but also of the nature of the depressive symptoms – primarily somatic versus primarily psychological. Factor, for example, analytic methods can reveal multidimensional structure but only (with the exception of certain methods not used in this study) at the expense of the assumption of response linearity. These issues have been developed in increasingly mathematical terms by McDonald (1965, 1981, 1982; McDonald & Ahlawat, 1974).

Unfortunately, Bech *et al* also skew the comparison between Rasch and factor-analytic methods by ignoring much psychometric research published in the last 20 years refining the use of factor-analytic methods: they apply the "eigenvalue greater than 1.0" criterion to decide the number of factors to rotate despite the fact that this has been shown to be a very bad indicator of the best number of important factors (e.g. Zwick & Velicer, 1986); they also use an arbitrary criterion to define the significance of factor loadings and apply no systematic indices of factor congruence.

This is not just an issue of parochial concern to psychometric hobbyists: multidimensional latent continuum structure is highly plausible in such data and the clarification of the extent to which such structure is, or is not, congruent across countries is an issue of very great importance as Bech *et al* note. A more extensive presentation of factor analytic results could considerably extend the insights Rasch analyses have given.

We would invite the authors to make their item data available to us to allow an analysis using factor-analytic methods. If, as we suspect, the factor-analytic methods do provide further insights into cross-cultural differences and consistencies we would submit a joint short communication to this *Journal* to complement the published excellent Rasch analyses and to redress some of the unfair polarisation against factor analysis in that publication.

MCDONALD, R. P. (1965) Difficulty factors and non-linear factor analysis. *British Journal of Mathematical and Statistical Psychology*, **18,** 11–23.
— (1981) The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, **34,** 100–117.
— (1982) Linear versus nonlinear models in item response theory. *Applied Psychological Measurement*, **6,** 379–396.
— & AHLAWAT, K. S. (1974) Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, **27,** 82–99.