endogenous or exogenous (e.g. amphetamine or bromocriptine), we may all be liable to develop psychotic symptoms. Thus a *sine qua non* of psychoses would be excessive dopamine, but a secondary susceptibility would also be required. Complementary research into the genetic and neurochemical aspects of such symptoms need not, therefore, be dissociated from allowing dopamine a central role in the generation of psychiatric illness.

T. H. TURNER
St Bartholomew's and Hackney Hospital
Homerton High Street
London E9 6BE

Reference

TURNER, T. H., COOKSON, J., WASS, J. A. H., DRURY, P. L., PRICE, P. A. & BESSER, G. M. (1984) Psychotic reactions during treatment of pituitary tumours with dopamine agonists. *British Medical Journal*, **289**, 1101–1103.

Imipramine Versus Phenelzine in Melancholias and Dysthymic Disorders

SIR: At a recent statistics seminar for students studying for Membership of the Royal College of Psychiatrists, the recent clinical trial by Vallejo *et al* (*Journal*, November 1987, **151**, 639–642) was discussed. It became apparent that the study was defective in a number of ways, so that the conclusions are difficult to support, and I feel it necessary to report some of the problems.

(i) The study is in fact two clinical trials, one for patients with melancholia and one for patients with dysthymic disorders. There were 32 patients in each trial. With this size sample, if 50% of patients improved on one drug, one would need a 95% improvement on the other to obtain a significant difference between the drugs at the 5% significance level with 80% power, giving a wide range in which to conclude that for imipramine and phenelzine 'patients responded equally well to both drugs'. In other words, the trial lacks power to conclude that the drugs were equivalent. This is clearly a case where confidence intervals should be given.

(ii) There is a statistical blunder in that the authors show that the variance of HRSD scores differs between imipramine and phenelzine (by 'Snedecor's test', which should have been referenced) and then proceed to compare means using the *t*-test. In fact, one of the assumptions underlying the validity of the *t*-test is that the variances are equal. Also, since the mean and standard deviation of the HRSD score are of similar size it is clear that the data are highly skewed. It would have been better to (a) give a graph of the data and (b) try a logarithmic transformation to see if this stabilised the variability. The graph would show the distribution of the data, and indicate whether their assumption about 'homogeneity' was valid. If the logarithmic transformation failed to stabilise the variance, a non-parametric test should be used.

(iii) It would have been better to compare changes in scores, rather than simply post-treatment values. Also, in view of the imbalance in the sexes between the two drugs in the melancholia group, an allowance for sex should have been made in the analysis.

It is now some years since White (1979) pointed out statistical errors in the *Journal*, but it is clear that there is still much room for improvement.

MICHAEL J. CAMPBELL
Dept of Medical Statistics and Computing
University of Southampton
Southampton General Hospital
Southampton SO9 4XY

Reference

WHITE, S. J. (1979) Statistical errors in papers in the British Journal of Psychiatry. *British Journal of Psychiatry*, **135**, 336–342.

SIR: We would like to make the following points regarding Dr Campbell's comments.

(i) In accordance with the norms of publication, it is not necessary to describe common statistical methods. In our case the interpretation of Snedecor's *F* does not lead to errors, given the context in which it appears.

(ii) Just as we indicated, the variance of the HRSD scores for the two major depression with melancholia groups are significantly different. But neither this fact nor the absence of normality in the distribution invalidates the use of Student's *t*-test. In fact, quite some time ago Bonneau (1960) demonstrated empirically that this test is extremely insensitive to the abnormality of the distribution and the heterogeneity of the variance when the *n* of the two groups is the same. This fact, added to the difficulty of interpreting the transformed scores, justifies not using them.

(iii) The use of non-parametric tests would reduce the power of the design.

(iv) Regarding power limits, it is important to point out that it is not the percentage of patients improved that is compared as Dr Campbell supposes but the difference in means in the HRSD score for both groups. By way of comparison, in the case of equality of variances and a 5% statistical significance