British Journal of Psychiatry (1985), **147**, 727–735

Correspondence

The Northwick Park ECT Trial

DEAR SIR,

The Clinical Research Centre, Division of Psychiatry (*Journal*, March 1984, **144**, 227–237) reported the results of a clinical trial which attempted to identify predictors of the response of depressed patients to real and simulated ECT. While this study is to be commended for its use of blind trials and a placebo ECT condition, it unfortunately contains a number of serious statistical errors. Given the implications of this study for clinical practice, I feel obliged to point out these errors, since they place nearly all the conclusions of the study under serious question.

The first such error the authors make is their stated willingness to accept non-significant results if they point to "clinically meaningful gradients". This places them on a very dangerous path from the outset, because while a skilled clinician may be able to derive a plausible clinical gradient from all but the most contradictory set of variables, this serves little purpose if that set of variables merely represents a random coincidence of events.

The results of their initial analysis of variance appear to indicate (clear details of the model used are not given) that none of the patient variables, i.e., age, sex, and presence of agitation, retardation or delusions had any effect on the therapeutic outcome of the ECT. The overall effect was also very weak, with real ECT showing a slight advantage over placebo ECT on only one of three measures (Johnstone *et al*, 1980). This difference was transient, disappearing between the end of treatment and a one-month follow-up. Nevertheless, the authors noted several non-significant trends in the data, namely the tendency of retarded and deluded patients to show a better response to real ECT than to placebo ECT.

It was on the basis of these trends that a discriminant function analysis was carried out using two predictive scales, to see if the response to real and placebo ECT was predicted by different groups of items. This is where apparently significant results were obtained; it is also where the most serious statistical errors have been made. No details of the discriminant analyses were given, in particular, significance levels for the functions themselves were omitted. Using each of the predictive scales, discriminant functions were obtained separately for the real and placebo conditions, to discriminate three levels of improvement. Different functions were obtained, which classified correctly the improvement of patients in the group from which they were derived, but not the other group. This applied to both predictive scales. It was concluded that the predictors of response to real ECT are distinct from those that determine response to placebo ECT, and that deluded depressed patients are more likely to respond to real ECT.

This conclusion is open to serious question. Firstly, the discriminant analysis used in this study incorporated stepwise inclusion of variables. This means that once the first variable has been entered into the function, it is unlikely that any other variable which is correlated with it will be entered for some time, since the next variable to be included is the one which provides the best discrimination *in combination* with the first. This means that even if the underlying causal variables of a function remain constant, for any given sample, different variables may be entered because of slight random variations in the percentage of discrimination they explain. So different discriminant functions do not necessarily imply different underlying causes: it also needs to be shown that the variables in the functions are unrelated, by a technique such as factor analysis or multiple correlation analysis.

Secondly, they have attempted to establish the accuracy of the discriminant functions, and their applicability to different groups, using classification error rates. This is not a valid procedure. As Gondek (1981) and Hand (1983) have recently pointed out, greatly inflated estimates of accuracy may result when classification functions are tested on the sample from which they are derived, particularly if the number of discriminating variables is large and the number of cases small. We have found that estimates of accuracy can be reduced from 80% to the chance level when the appropriate validation techniques are used. This may explain why the discriminant functions obtained from the real ECT group did not correctly classify improvement by the placebo ECT group, and vice-versa.

There are a number of techniques which can be used to overcome this problem. Cross-validation can be employed. The discriminant functions are generated on one part of the sample and then applied to another part. This can take the form of a split-half reliability test, or, if the overall sample size is small, a smaller percentage can be excluded and the reliability test run a number of times (Frank, Massey & Morrison, 1965). In the extreme case, only

https://doi.org/10.1192/bjp.147.6.727 Published online by Cambridge University Press

one case is excluded at a time and the number of tests is equal to the number of cases. This is the "leaving one out" or "jacknife" method (Lachenbruch & Mickey, 1968). Alternatively one can generate a random sample with the same number of variables and cases, and estimate the random component of the classification function to use as the base-line for significance tests. Another possibility is to use "bootstrap" methods, where the sample is duplicated a number of times and then a number of random samples are drawn to gain an estimate of the bias present in any given sample of the same size (Efron, 1982; McLachlan, 1980).

I would urge the Clinical Research Centre to carry out a re-analysis of their data, and to publish their revised findings. In the meantime, clinicians should set aside their recommendation that ECT is an appropriate treatment for delusional depression; it may turn out that this otherwise well designed study indicates that the effect of ECT is almost wholly a placebo effect.

GEOFFREY W. STUART

Health Commission of Victoria,
Mental Health Research Institute,
35–37 Poplar Road,,
Parkville, Victoria 3052,
Australia

References

EFRON, B. (1982) *The Jacknife, the Bootstrap and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics.

FRANK, R. E., MASSEY, W. F. & MORRISON, D. G. (1965) Bias in multiple discriminant analysis. *Journal of Marketing Research,* **2**, 250–258.

GONDEK, P. C. (1981) What you see may not be what you get: Discriminant analysis in statistical packages. *Educational and Psychological Measurement,* **41**, 267–281.

HAND, D. J. (1983) Common errors in data analysis: The apparent error rate of classification rules. *Psychological Medicine,* **13**, 201–203.

JOHNSTONE, E. C., DEAKIN, J. F. W., LAWLER, P., FRITH, C. D., STEVENS, M., MCPHERSON, K. & CROW, T. J. (1980) The Northwick Park ECT Trial. *Lancet, ii,* 1317–1320.

LACHENBRUCH, P. A. & MICKEY, M. R. (1968) Estimation of error rates in discriminant analysis. *Techometrics,* **10**, 1–11.

MCLACHLAN, G. J. (1980) The efficiency of Efron's "Bootstrap" approach applied to error rate estimation in discriminant analysis. *Journal of Statistical Computation and Simulation,* **11**, 273–279.

Dr Crow and Colleagues Reply

DEAR SIR,

Dr Stuart claims to have detected "serious statistical errors" in the analysis of the Northwick Park ECT trial. Some of his assertions are based upon a failure to separate our original report of the overall therapeutic effect (Johnstone *et al*, 1980), which we regard as the most reliable information we have of the magnitude of the therapeutic effect, from our later analysis (*Journal*, March 1984, **144**, 227–237) of possible predictors of response to real ECT. The conclusions of the latter paper are much less certain as will be apparent from the summary which indicates that "The limited size of the sample does not allow firm conclusions."

Concerning the former paper Dr Stuart writes that "the overall effect was very weak". The size of the effect is apparent in the figure in our original paper and the difference between the groups over the course of the four-week trial was significant at the 1% level. None of Dr Stuart's criticisms appear to be relevant to this conclusion although this may not be apparent to the casual reader of his letter.

Dr Stuart's criticisms are, however, relevant to our analysis of potential predictors of ECT response. Here we are surprised that Dr Stuart regards our paper as having made anything more than tentative suggestions concerning predictors. Thus in addition to the above statement in the summary, the Discussion includes the following:

"Extensive examination of this data did not show clear-cut predictors of response to real or simulated ECT, but this was not unexpected in view of the relatively small sample size". Our main conclusions concerning predictors are included in the subsequent sentences:

"Firstly: we found little support for our previous suggestion . . . that the predictors of response to ECT are merely the predictors of satisfactory response to treatment or even a generalised tendency to satisfactory outcome. . . .

Secondly, the analysis does not support the view that a predominance of endogenous features is a specific predictor of a response to real ECT. The most salient predictor of response to real ECT is probably the presence of delusions. . . . Retardation may be relevant but it may also be associated with response to simulated ECT."

The burden of Dr Stuart's letter is that a different method of statistical analysis would have given a different and more certain conclusion.

We doubt that this is the case but think rather that there is an overriding limitation in the size of our sample. Our sample was primarily collected to test for an overall therapeutic effect of ECT, not to test for interactions. No amount of statistical sophistication can get from a study more information than is contained in the data originally included. Our sample ($n = 62$ completors) took 3½ years to accumulate but as a basis for prediction of differen-

https://doi.org/10.1192/bjp.147.6.727 Published online by Cambridge University Press