Brit. J. Psychiat. (1966), **112**, 101–106

Correspondence

THE DIAGNOSIS OF DEPRESSIVE SYNDROMES AND THE PREDICTION OF E.C.T. RESPONSES

Dear Sir,

In their paper (*Journal*, August 1965, page 659), Carney, Roth and Garside describe a sophisticated factor analysis of diagnostic items and treatment-response which fortifies their opinion that endogenous depression is qualitatively distinct from neurotic depression. It would be agreed, I am sure, that, in any research, formidable techniques of statistical analysis cannot reveal truths impossible of access by the method of data collection used.

The data collected in the Newcastle study were subjectively-determined scores. Before treatment the patients were assessed "by one of the authors" who "found it necessary to supplement the information from the interview with material from the case notes and the observations of the nursing staff". Subsequently, "follow-up assessments were made at 3 months and 6 months by one of the authors", by clinical examination and "consultation with medical staff and charge nurses". In other words, there was always an element of indecision.

We may fairly suppose that when the authors made their assessments they were already of the opinion that endogenous and neurotic depression were distinct entities. The analysis reported is an analysis of their opinions, including their opinions of improvement after E.C.T. Did the analysis reveal something that was really present in the patients? Or did it reveal primarily what was present in the authors' minds, through which bimodally-biased filters all the information had had to pass before reaching the man-made computer?

The response-to-E.C.T. argument must be rejected, since the "response" was actually a score subjectively determined by authors who, at the time, knew the history and also whether E.C.T. had been given. We are left only with the following answer to possible observer preconceptions: "It would, however, be difficult to explain along these lines the close correspondence between the correlation of features with diagnosis, on the one hand, and their loadings for the bipolar factor on the other." Would it? Or does one, not without admiration, descry here some rather splendid statisticsmanship?

I should acknowledge my personal sympathy with the authors' belief concerning the bimodality of depressive disorder. I am, however, beset by nagging doubts about the evidence. Had the authors employed an assessor from among those psychiatrists who firmly believe that endogenous and neurotic depressions are not distinct, and the factor analysis had yet produced the same results, I should have been convinced. Perhaps surprised too.

Ian Oswald.

University Department of Psychiatry,
Royal Perth Hospital,
Perth,
Western Australia.

Dear Sir,

At first sight the article by Carney *et al.* (August 1965) contains an impressive mass of statistical data. However, when one looks more closely and critically at these statistics, some serious fallacies can be detected.

A number of clinical features of depressive illnesses are analysed, and correlation coefficients are published which have been worked out to three decimal places by a computer. Unfortunately many of these clinical features cannot be expressed in numerical terms; examples are "hysterical features or attitude", "hopeful attitude towards illness", "constipation", and "adequate personality". The authors make the following statement about their method of assessment of these features: "A score of one was assigned to each clinical feature if present, and a score of nought if absent, except in the case of 'guilt' when delusions scored two, 'feelings', one, and 'guilt-free' nought."

The method of assessment was therefore quite arbitrary, and it seems to me illogical that correlation coefficients worked out to three decimal places should be derived from these original rough assessments. Clearly, many of the clinical features examined cannot be said to be either present or entirely absent, and the application of the most elaborate statistical techniques available cannot alter the fact that many of the basic data cannot be expressed quantitatively at all, or at best can only be assessed very roughly, as is illustrated by the authors' rating scale for "guilt".

The article illustrates well an ever-increasing tendency in published work on psychiatric subjects to substitute for clinical observation and description a pseudo-scientific mathematical or statistical approach. I suppose it all began with the idea that "intelligence" can be accurately measured and expressed in mathematical terms.

W. J. STANLEY.

15 St. John Street,
Manchester.

DEAR SIR,

Professor Oswald's first point concerns our method of data collection. He complains that the data were "subjectively-determined scores" and that "an element of indecision" was always present. If this is meant as a criticism, it is a criticism that embraces nearly all attempts at ordinary clinical diagnosis in psychiatry, which is inevitably based on subjectively-determined (as opposed to physically measurable) data. This does not at all invalidate the use of statistical methods, provided that the errors of judgment are randomly distributed. It would indeed be unfortunate, and erroneous, if the "indecisiveness" of clinical judgment came to be regarded as *ipso facto* invalidating the use of statistical methods in psychiatry, since the opposite is true. Statistics can be especially useful under just these circumstances, precisely because unreliability always *reduces* correlations and thus statistically significant conclusions cannot be produced by unreliability of data. Any correlations found are despite and not because of the unreliability of the individual clinical ratings. This is also the answer to one of Dr. Stanley's criticisms.

Professor Oswald's next point concerns *bias*, which is not the same as unreliability. He suggests that bias may have been operating because (i) the data could have been recorded in such a way as to fit in with our preconceived notions of neurotic and endogenous types of depression; and (ii) the recorded state of follow-up could have been influenced by knowledge of the previous history.

With regard to (i), it is really very difficult to understand just how this could have brought about the results obtained. Of course we believed that neurotic and endogenous depression were distinct entities, since previous work in the department had led us to this belief. The 35 items selected for analysis were indeed chosen precisely because of their supposed discriminating function. Yet the factor loadings on the bipolar factor did not fall into two discrete groups; on the contrary, there were nine items in the intermediate range, with loadings less than $\pm 0 \cdot 300$. Among these was the item "worse in morning", which was certainly expected to have a high correlation with diagnosis, but in fact did not (correlation $+0 \cdot 143$). One wonders just at what point the rater's bias began to operate.

Secondly, the correlation of each item with diagnosis agreed remarkably well with its factor loading on the bipolar factor. To arrive at such a close fit by bias alone would have demanded extraordinary mathematical insight on the part of the rater; the data would have to have been recorded in such a way that the 595 correlations between the 35 features produced loadings proportional to the correlations with diagnosis. Professor Oswald suggests that the fit was obtained by "splendid statisticsmanship". In fact, the loadings were derived directly from the computer without rotation or other manipulation.

Professor Oswald finds a second source of bias in the follow-up ratings. He says: "The response-to-E.C.T. argument must be rejected since the 'response' was actually a score subjectively determined by authors who, at the time, knew the history and also whether E.C.T. had been given." Actually, *all* patients had been given E.C.T., and this knowledge could not have influenced the rating. Moreover, the rater made his follow-up assessments without reference to the original diagnoses. To the extent that he might have remembered some of the patients, it would have been methodologically desirable to have employed an independent rater; considerations of manpower made this impracticable.

Professor Oswald suggests that had an assessor from another school been employed, and the same results obtained, he would have been convinced. But would he? He could argue that unconscious biasses operated, that the assessor changed his allegiance during the course of the investigation, and so on. The appropriate action on his part would be to carry out a similar investigation himself or to await the publication of the results of other investigators (for example, Sandifer, Wilson and Green, *Amer. J. Psychiat.*, in press). Additional studies are also under way at the Massachusetts Mental Health Center by Rosenthal and Klerman, and in the National Institute of Mental Health Psychopharmacology Service Center Collaborative Depression Project. We would, however, stress the need for careful definition of the prognostic features under investigation, and also suggest that the use of standardized, structured interviewing techniques should be further explored.

Dr. Stanley complains that the assessments were too rough to be treated by statistical methods. As already stated, any errors in assessment may be expected to balance out, provided that they are