CLASSIFICATION OF THE DEPRESSIVE DISORDERS
BY NUMERICAL TYPOLOGY*

JOSEPH L. FLEISS
Biometrics Research, New York State Department of Mental Hygiene,
722 West 168th Street, New York, N.Y. 10032

(Received 13 April 1971)
(Revised 28 July 1971)

INTRODUCTION

One of the liveliest debates in psychiatric nosology today concerns the nature of depressive disorders. There exist, on the one hand, studies which indicate the existence of two⁴⁻⁴, three⁵ and even four⁶⁻⁷ distinct subtypes of depression. There exist, on the other hand, studies which have failed to indicate the existence of more than a single type.⁸⁻¹⁰

The issue is not a trivial one. If different subtypes really exist, the study of patients assigned to them can have important consequences in identifying etiological factors producing the various types of depression, in predicting the future clinical course of these types, and in specifying the best treatment for patients from the various types. If, however, there is but a single continuous type, then more information on etiology, prognosis and treatment is provided by a patient’s position on that continuum than by his assignment to one of a number of arbitrary categories.

A thorough review of the controversy is given by Kendell.¹¹ In this paper a critique is offered of some of the techniques which have been used to derive or test numerical typologies of depression. A statistical model which hypothesizes two overlapping normal distributions is then applied to data on a sample of 104 depressives. A mixture of two types just barely provides a significantly better fit to the data than does a single type.

A CRITIQUE OF METHODS FOR IDENTIFYING TYPES

Finding multimodal distributions

The strongest statistical evidence for the existence of subtypes within a population would appear to be the presence of two or more modes in a plotted frequency distribution. This was the kind of evidence put forth, for example, by Carney et al.² and by Sandifer et al.³ to support the contention that there were two subtypes of depression. Such findings of multimodality should not, however, be taken at face value, especially when the distributions

* This work was supported in part by grant MH-09191, Bilateral Study of Diagnosis of the Mental Disorders, from the National Institute of Mental Health.
are derived from psychiatric or psychological ratings. Kendall\textsuperscript{11} and Paykel \textit{et al.}\textsuperscript{12} have pointed out how easily biases or preconceptions on the part of the raters can yield bimodal distributions. When, as in the study of Carney \textit{et al.},\textsuperscript{2} ratings are made by psychiatrists committed to the model of two types of depression, and furthermore are made only after an interview during which the rater might well have made up his mind regarding the type to which the patient belonged, the demonstration of bimodality can hardly be considered a definitive proof of the existence of two types.

\textit{Inverted factor analysis}

A number of the other reports which have indicated the possible existence of different subtypes of depression have relied on one or another of the many methods of cluster analysis which currently exist. These cluster analytic methods have in common the attempt to constitute homogeneous groups on the basis of measured likenesses among the subjects or patients to be classified. Probably the most popular of these methods in psychiatry is inverted, or $Q$ factor analysis. This clustering technique was employed, for example, by Overall \textit{et al.}\textsuperscript{5} in their study of depression.

Suppose that each of a sample of subjects is measured on each of a number of variables. The first step in $Q$ factor analysis is to scale each variable to a common mean (usually zero) and a common standard deviation (usually unity) across the entire sample of subjects. The second step is to obtain a product--moment correlation coefficient, the so-called $Q$ correlation, for each pair of subjects. This unusual coefficient requires some description.

Each subject has a set of scores, in standard units, on the measured variables. For each pair of subjects, then, there exist two such sets, one from each subject. The $Q$ correlation is then simply the usual product--moment correlation coefficient between these two sets of scores. Its magnitude describes, not the degree of association between two variables, but rather the extent of similarity between two subjects.

Figure 1 presents the standard method for representing pictorially subjects' sets of scores, or profiles. Since each score for subject II is one unit higher than the corresponding score for subject I, their profiles are parallel. To many researchers, it is the similarity or dissimilarity of the shapes of two profiles, and not any possible difference in mean levels, which contributes to their likeness. Thus, the profiles for subjects I and II would be judged to be perfectly alike, and the $Q$ correlation reflects this perfect likeness by assuming the value one.

Figure 2, on the other hand, shows two profiles for which the peaks and valleys of one do not correspond to the peaks and valleys of the other. Even though the profiles are at the same mean level, many investigators would conclude that they are not at all similar. The $Q$ correlation reflects this lack of similarity by assuming the value zero.

The final step in $Q$ analysis is to subject the matrix of all correlations among subjects to a factor analysis. The number of factors found is taken to be the number of types which exist, and each type is constituted by assigning to it those subjects whose largest factor loadings are on the corresponding factor.

A number of criticisms have been leveled against this method for numerically deriving types.\textsuperscript{13-16} Some of the more cogent criticisms are that correlations among the variables, in the usual sense, are ignored in calculating $Q$ correlations; that the maximum number of
types which may be found by inverted factor analysis is severely restricted by the number of variables which are measured; and that the $Q$ correlation, in addition to being virtually uninterpretable, is actually a poor measure of similarity.

**Fig. 1.** Two parallel profiles with a $Q$-correlation of unity.

**Fig. 2.** Two dissimilar profiles with a $Q$-correlation of zero.
The point has already been made that it is the degree of parallelism of two profiles which should determine, according to some investigators, the magnitude of the measure of similarity. It should therefore follow that two profiles which are not parallel should be measured as being less than perfectly alike. The $Q$ correlation does not possess this property. Consider the two profiles in Fig. 3. Each score in profile II is twice the corresponding score in profile I, plus one. The two profiles are clearly not parallel, but yet, because of the perfect linear relation between them, their $Q$ correlation is unity. Because the $Q$ correlation measures nonparallel profiles as being perfectly similar, the results of inverted factor analyses must be suspect.

![Diagram](image)

**Fig. 3.** Two nonparallel profiles with a $Q$-correlation of unity.

*The Friedman–Rubin technique*

A clustering method which seems to be free of the above criticisms is due to Friedman and Rubin. Their technique has been applied to the study of depression by Paykel. The Friedman–Rubin method shares a defect with almost every existing clustering technique, however: there is provided no criterion for deciding that the sample is a homogeneous entity to begin with, and therefore that no clustering should be attempted.

Forgy assessed the performance of a number of clustering methods when applied to samples generated according to various models. When the samples were mixtures of well separated clusters, most of the methods succeeded in recapturing the clusters. These methods
thus possess a property analogous to a low probability of a type II error for a statistical test: when subtypes clearly exist, these methods have a high likelihood of finding them.

When the samples were homogeneous, on the other hand, i.e. generated by sampling from a single multivariate normal distribution, FORGY found that these methods nevertheless detected presumed subtypes. Further, the types found by one method often differed radically from those found by another. Again in analogy with statistical tests, it appears that the low probability of a type II error is bought at the price of a high probability of a type I error: even when subtypes do not exist, these methods will, with a high likelihood, find them.

The FRIEDMAN–RUBIN method was not one of those assessed by FORGY, nor has it, to the author's knowledge, been so studied by anyone else. In their report of their method, FRIEDMAN and RUBIN\textsuperscript{17} described its success in recapturing what were known to be distinct subtypes. One cannot yet tell, however, how well or poorly it performs when applied to a single homogeneous sample.

Overly stringent criteria for subtypes

At this point a definition is required of what constitutes a homogeneous and what a heterogeneous group. There seems to be no way to avoid introducing parametric assumptions into the definition. A group will be said to be homogeneous if the joint distribution of the observations on the subjects within it is a multivariate normal distribution. A group will be said to be heterogeneous if the joint distribution is a mixture of multivariate normal distributions.

The advantage of such a definition is that the sample may be tested for homogeneity before any attempt at clustering is made. A disadvantage, clearly, is that the assumed form for the joint distribution may be totally incorrect.

A criterion something like the above was in effect adopted by ROSENTHAL and GUDEMAN,\textsuperscript{19} KENDALL\textsuperscript{9,11} and KENDALL and GOURLEY.\textsuperscript{10} A number of samples of depressed patients were represented in their studies, with unimodal distributions almost invariably being found. Since multimodality would be a clear indicator of the existence of subtypes, these authors took its absence as an indication that their data did not confirm the hypothesis of subtypes.

The implicit criterion of these authors is that only the finding of multimodal distributions can confirm the hypothesis that subtypes exist. This criterion is much too stringent. It should be borne in mind that, whereas multimodality is a sufficient indicator of the existence of subtypes, it is not a necessary one. EISENBERGER\textsuperscript{20} and BEHBOODIAN\textsuperscript{21} have shown that two mixed normal distributions will result in a bimodal distribution only if, roughly speaking, the means of the two distributions are well separated and the proportions in which the two distributions are mixed are nearly equal. Thus, if the means are fairly close, or if one distribution is present in a much greater proportion than the other, the mixed distribution will be unimodal rather than bimodal.

A criterion related to unimodality which has sometimes been adopted\textsuperscript{10,12} is the insignificant magnitude of the chi square statistic for testing the goodness of fit of a normal distribution. The defect in this criterion is that the classic chi square test is not sufficiently sensitive to the kinds of alternative hypotheses which are most reasonable in this problem,
namely, alternatives which specify a mixture of distributions. The test for the goodness of fit of a single normal distribution may fail to reject the hypothesis, indicating that such a distribution is consistent with the sample. This does not mean, however, that a mixture of normal distributions would not provide a significantly better fit.

Even the rejection of the hypothesis of a single normal distribution has not always been adequately followed up; e.g. Paykel et al.\textsuperscript{12} found significant departures from normality but ascribed the departures more to skewness than to a mixture of different normal distributions. What was ignored here was the fact that two or more mixed normal distributions may yield a unimodal skewed distribution. Indeed, successful analyses into two or three normal populations of just such distributions have been reported.\textsuperscript{22}

An improved criterion for subtypes

A criterion which seems to provide a better chance of detecting subtypes than either multimodality or the significance of chi square is the following. Suppose that each of \( N \) subjects is scored on a single variable, or is given a single score which summarizes his values on a number of variables. Let \( x_i \) denote the score for the \( i \)th subject. Since the maximum likelihood estimates of the parameters of a single normal distribution are

\[
m = \frac{1}{N} \sum_{i=1}^{N} x_i / N \quad \text{and} \quad s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m)^2 / N,
\]

the maximum value of the logarithm of the likelihood for a single normal distribution is

\[
L_1 = -\frac{N}{2} \left( 1 + \ln (2\pi) + \ln (s^2) \right). \tag{1}
\]

The joint density function for a mixture of two normal distributions with equal variances is

\[
f (x_1, \ldots, x_N) = (2\pi \sigma^2)^{-N} \prod_{i=1}^{N} \left\{ P \cdot \exp \left( -(x_i - \mu_1)^2 / 2\sigma^2 \right) 
+ (1 - P) \cdot \exp \left( -(x_i - \mu_2)^2 / 2\sigma^2 \right) \right\}. \tag{2}
\]

If \( p, m_1, m_2 \) and \( \nu^2 \) maximize \( f \) (these maximum likelihood estimates of \( P, \mu_1, \mu_2 \) and \( \sigma^2 \) must be found iteratively), then the maximum value of the logarithm of the likelihood for two mixed normal distributions is

\[
L_2 = -\frac{N}{2} \left( \ln (2\pi) + \ln (\nu^2) \right) + \sum_{i=1}^{N} \ln \left\{ P \cdot \exp \left( -(x_i - m_1)^2 / 2\nu^2 \right) 
+ (1 - P) \cdot \exp \left( -(x_i - m_2)^2 / 2\nu^2 \right) \right\}. \tag{3}
\]

The same variance \( \sigma^2 \) had to be assumed for the two mixed distributions because otherwise the maximum likelihood estimates are indeterminate.\textsuperscript{23}
If the hypothesis of a single normal distribution is true, then, if \( N \) is large, the quantity 
\[
U = 2(L_2 - L_1)
\]
will be approximately distributed as chi square with two degrees of freedom.\(^{24}\) The degrees of freedom are two because the model of two mixed normal distributions specifies four independent parameters whereas the model of a single normal distribution specifies two. The degrees of freedom equal the difference between these numbers of parameters.

If \( U \) is small, say if \( U < 5.99 \), the critical value of chi square with two degrees of freedom when the significance level is 0.05, the inference may be drawn that a single normal distribution is sufficient to describe the sample. If \( U \) is large, one may proceed either to fit a mixture of three normal distributions or, if one is unwilling to entertain the possibility of more than two distributions, to assign the subjects to one or the other of the two types which were found. This assignment may be made by the specialization to a single variable of the classic method for classifying observations into one of two normal distributions.\(^{25}\) The method reduces to the following rule, where \( x \) represents the score for the subject to be assigned, and where it is assumed that \( m_2 > m_1 \):

\[
\text{If } x < \frac{m_1 + m_2}{2} + \frac{1}{m_2 - m_1} \ln \frac{p}{1 - p}, \text{ assign the subject to the first type; otherwise, assign him to the second.}
\]

**Fitting a Mixture of Normal Distributions to a Sample of Depressives**

*Description of sample*

As part of a study comparing psychiatric diagnosis in the United States and United Kingdom, a total of 250 consecutive admissions to the inpatient service at a London public mental hospital and an equal number in a New York State mental hospital were intensively studied by a cross-national team of psychiatrists. Details of the study design, of the methods of evaluation, and of the results are reported elsewhere.\(^{26}\)

Here, only data from the London sample are presented. All patients were aged between 20 and 59 years. In addition to age, the only other criterion for inclusion in the study was that the patient be a new admission to the hospital (not a transfer or a return from leave).

As soon as possible after admission (usually within 48 hr of admission and rarely more than 72 hr later), each patient was interviewed by one of the project psychiatrists using a standardized mental state schedule.\(^{27}\) Soon after, both the patient and a close informant were interviewed using a less standardized history schedule. Using the information from the mental state and history interviews, the project psychiatrist who had conducted the interviews made a diagnosis using the eighth edition of the International Classification of Diseases.

Of the 250 patients in the London sample, 104 were given a primary diagnosis of depression by the project psychiatrists. Thirty-nine were diagnosed neurotic depression and 65, psychotic depression. Included in the latter category were involuotional melancholia, the depressed type of manic-depressive psychosis, and psychotic depressive reaction. Some 70 per cent of the London depressives, as opposed to some 60 per cent of the total London sample, were females.
The measurement of psychopathology

Each of these patients was scored on 24 dimensions of psychopathology (see Table 1) found by a factor analysis of nearly 700 items describing the patient's current mental state. The depression factor measured, *inter alia*, loss of interests, inability to concentrate, lack of energy, suicidal thoughts and diminished appetite. Early morning wakening, depressive dreams, delusions of guilt and other depressive delusions and hallucinations failed to contribute to the depression factor because of low correlations with the other behaviors which it measured.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Psychotic depressives</th>
<th>Neurotic depressives</th>
<th>Coefficient in discriminant function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Depression</td>
<td>26.37</td>
<td>11.21</td>
<td>23.72</td>
</tr>
<tr>
<td>Phobic anxiety</td>
<td>4.32</td>
<td>4.53</td>
<td>5.51</td>
</tr>
<tr>
<td>Reported restlessness</td>
<td>1.98</td>
<td>1.80</td>
<td>1.90</td>
</tr>
<tr>
<td>Observed restlessness</td>
<td>0.77</td>
<td>1.53</td>
<td>0.15</td>
</tr>
<tr>
<td>Retarded speech</td>
<td>0.72</td>
<td>1.91</td>
<td>0.18</td>
</tr>
<tr>
<td>Retarded movement</td>
<td>0.82</td>
<td>1.37</td>
<td>0.26</td>
</tr>
<tr>
<td>Mania</td>
<td>0.08</td>
<td>0.41</td>
<td>0.05</td>
</tr>
<tr>
<td>Somatic concerns</td>
<td>0.68</td>
<td>1.42</td>
<td>0.15</td>
</tr>
<tr>
<td>Observed belligerence</td>
<td>0.35</td>
<td>1.16</td>
<td>0.13</td>
</tr>
<tr>
<td>Reported belligerence</td>
<td>1.14</td>
<td>1.57</td>
<td>1.97</td>
</tr>
<tr>
<td>Obsessions</td>
<td>0.62</td>
<td>1.42</td>
<td>0.44</td>
</tr>
<tr>
<td>Self neglect</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Disorientation</td>
<td>0.18</td>
<td>0.68</td>
<td>0.36</td>
</tr>
<tr>
<td>Lack of insight</td>
<td>0.68</td>
<td>1.15</td>
<td>0.44</td>
</tr>
<tr>
<td>Depersonalization</td>
<td>1.02</td>
<td>2.58</td>
<td>0.44</td>
</tr>
<tr>
<td>Paranoid delusions</td>
<td>0.42</td>
<td>1.37</td>
<td>0</td>
</tr>
<tr>
<td>Grandiose delusions</td>
<td>0.17</td>
<td>1.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Control delusions</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Visual hallucinations</td>
<td>0.12</td>
<td>0.88</td>
<td>0.02</td>
</tr>
<tr>
<td>Auditory hallucinations</td>
<td>0.83</td>
<td>2.90</td>
<td>0.38</td>
</tr>
<tr>
<td>Bizarre behavior</td>
<td>0.09</td>
<td>0.42</td>
<td>0</td>
</tr>
<tr>
<td>Nonsocial speech</td>
<td>0.09</td>
<td>0.52</td>
<td>0.02</td>
</tr>
<tr>
<td>Flat affect</td>
<td>0.28</td>
<td>1.12</td>
<td>0.08</td>
</tr>
<tr>
<td>Incomprehensibility</td>
<td>0.40</td>
<td>1.21</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The phobic anxiety factor was measured by the presence of specific situations giving rise to anxiety and by physiological concomitants of anxiety. Observed restlessness included such observed behaviors as marked agitation, fidgeting and pacing; retarded movement included such observed behaviors as a slowness of movement and shuffling gait; and somatic concerns included such behaviors as preoccupation with physical complaints and the mistaken belief that an organ system was diseased.

As a means of generating a distribution of scores which would be amenable to a parametric mixture analysis, a discriminant analysis was applied to the sample of 104 depressives, seeking that linear combination of the factors which would best distinguish the 39 neurotic depressives from the 65 psychotic depressives. Only the factors for which the $t$ ratio for the
difference between the neurotic and psychotic depressives was greater than one were included in the discriminant analysis.

Table 1 gives, for each factor, the means and standard deviations for the psychotic and neurotic depressives, the corresponding $t$ ratio, and the coefficient for that factor in the linear discriminant function if its $t$ ratio exceeded unity. The neurotic depressives scored significantly higher than the psychotic depressives on only one factor, reported belligerence (including such reported behaviors as having fits of anger and having hit or attacked people). The psychotic depressives scored significantly higher than the neurotic depressives on three factors: observed restlessness, retarded movement and somatic concerns.

**Fitting distributions**

Each of the 104 depressives was given a score on the discriminant function by applying the coefficients in the rightmost column of Table 1 to his set of factor scores. The resulting distribution of discriminant function scores is shown in Fig. 4, and is seen to be unimodal but skewed to the right. Superimposed on the sample histogram is the frequency curve of the best fitting normal distribution, with a mean of 0.81 and a standard deviation of 1.34. For testing the goodness of fit of this distribution, intervals at the extremes were combined to give a total of 11 to which the classic chi square was applied. The value of chi square was 13.86, which with eight degrees of freedom was significant only at the 0.10 level.

![Histogram](https://via.placeholder.com/150)

**Fig. 4.** Observed frequency distribution on discriminant function scores for 104 depressives, best fitting normal distribution, and best fitting mixture of two normal distributions.

The data are thus consistent with a single normal distribution, although this is no proof that the data would not be significantly more consistent with a mixture of two normal distributions. To test this possibility, such a mixture was fitted by maximum likelihood to
the ungrouped frequency distribution. The estimated proportions of mixture were 0.87 and 0.13, the estimated means were 0.47 and 3.05, and the estimated standard deviation was 1.02. This mixture also appears in Fig. 4, and seems to account for the skewness observed in the sample histogram better than a single normal distribution. To check that these estimates provided the absolute and not merely a local maximum, the likelihood function was evaluated at many other parameter points but never increased in magnitude.

The required mathematics was quite complex. The initial estimates of the four parameters of the mixture were obtained so that they reproduced the first four sample moments. Even estimation by the method of moments is complicated.\(^{28}\) The computer program for proceeding iteratively from the initial estimates to those maximizing the likelihood was a modification of the pioneering maximum likelihood programs written by \textsc{Wolfe}.\(^{24}\)

The natural logarithm of the likelihood of the best fitting single normal distribution was \(L_1 = -82.79\) (see equation 1), and that of the best fitting mixture was \(L_2 = -79.68\) (see equation 3). The value of the approximate chi square statistic, with two degrees of freedom, is \(U = 2(L_2 - L_1) = 6.22\). This value is barely significant at the 0.05 level.

\section*{DISCUSSION}

This exercise has neither proved nor disproved that depression is a single homogeneous entity with respect to psychopathology. One can, in general, claim to have proved a statistical hypothesis only if one has failed to reject the hypothesis with data on a large sample. In the case of testing for normal mixtures, a large sample is one numbered in the thousands.\(^{29}\) The sample of 104 depressives is not only not large, it is puny. Conversely, with respect to an issue as important as the classification of psychiatric disorders, one should reject the null hypothesis of a single type only if the significance criterion is stringent, say a probability of 0.005 or 0.001. The probability level found above, 0.05, is hardly small enough to permit the inference that there are two types of depressives.

What this exercise has shown, rather, is the feasibility of a method of data collection and analysis which should provide a definitive answer to the question of a dichotomy between neurotic and psychotic depression. The reliance on a structured interview protocol aimed at providing a complete picture of psychopathology, and not just of depressive symptomatology, served to reduce the possibility of halo effects producing a spuriously bimodal curve. The interviewing of a cross-section of patients, and not just of depressives, served to reduce the possibility of a confounding of the diagnosis of neurotic or psychotic depression with the evaluation of psychopathology.

The strategy of working with scores on the discriminant function distinguishing neurotic from psychotic depressives was adopted because it seemed to provide the best chance of picking up a mixture if one was there to be found. \textsc{Paykel}'s report,\(^{7}\) the most recent to indicate that subtypes of depression are identifiable, showed that the distinction is essentially between neurotic and psychotic depression. \textsc{Paykel}, it is true, found four types, but three seemed to consist of neurotics whereas one consisted of psychotics. Thus, if the distribution on a discriminant function fails to be separable into two normal distributions, it is hard to imagine a distribution on any other summarization of mental state factors which would be so separable. In addition, scores on a discriminant function, being linear combinations
of a number of variables, should be approximately normally distributed if the number of variables is fairly large and if the variables are no more than moderately correlated. This makes the normality assumption underlying the mixture analysis a realistic one.

Recent research\textsuperscript{30,31} has indicated that the distinction between unipolar depression (depression alone in the history) and bipolar depression (depression plus mania in the history) may be more valid than that between neurotic and psychotic depression. The evidence is not yet extensive enough, however, to be considered definitive.

\textit{The exclusion of psychiatric history data}

The decision was deliberate to base the analysis on mental state psychopathology alone rather than on its combination with measures of basic personality or with such features of past psychiatric history as the suddenness of the onset of illness, the presence of precipitating events, the existence and kind of previous episodes, etc. One reason is that the reliability of personality and of psychiatric history data has been of a much smaller order of magnitude than that of mental state data,\textsuperscript{6,12,32} although the reliability is improving.\textsuperscript{30,33,34}

A second reason is that, whereas features from the history are essential to the distinction between the subtypes of depression identified as reactive and endogenous, they are not as essential to the distinction between the subtypes identified as neurotic and psychotic depression. There is some overlap between these two ways of splitting depression, but Paykel's work\textsuperscript{7} indicates that the latter partition, the one definable mainly in terms of mental state psychopathology, is a more useful one than the former. The inclusion in the analysis of history variables would therefore have meant the incorporation of unreliable and possibly irrelevant information.

None of this is to imply that measures other than those of psychopathology will not ultimately prove useful in the partitioning of samples of depressives. Whatever these other measures prove to be—of personality, historical or biochemical variables—their reliability will have to be improved beyond what it is now.

\textbf{SUMMARY}

Following a critique of some of the methods used in previous studies of the typology of depression, data on a sample of 104 hospitalized depressives were subjected to a parametric mixture analysis. The results were equivocal, inasmuch as a mixture of two types provided a significantly better fit than a single type at only the 0.05 level.

It appears that the conclusive statistical evidence for the existence of sub-types of depression will have to be the demonstration, using reliable data, that depression is better described by a mixture of distributions than by a single homogeneous distribution. Sample sizes much larger than any so far studied, and the measurement of historical, personality or other variables with precision approaching that of the measurement of psychopathology, are necessary for a definitive conclusion.

\textit{Acknowledgements}—Critical comments on an earlier version of this paper by Dr. J. Zubin, Director of the Bilateral Study; Dr. O. J. Dunn of U.C.L.A.; Dr. R. E. Kendell of the Institute of Psychiatry, London; Dr. P. A. Lachenbruch of the University of North Carolina; and Dr. E. S. Paykel of Yale University are gratefully acknowledged.

The mental state ratings were provided by Drs. J. E. Cooper, J. R. M. Copeland, T. Farkas, B. J. Gurland, R. E. Kendell, N. Sartorius and L. Sharpe of the Bilateral Study.
REFERENCES


