Cluster Analysis and the Classification of Depression

By NANCY C. ANDREASEN, WILLIAM M. GROVE and RALPH MAURER

During recent years there has been considerable controversy concerning the classification of depressive disorders. Investigators disagree as to whether depressive disorders are a single phenomenon with varying degrees of severity or are a group of two or more discrete illnesses with differing courses, prognoses, responses to treatment, familial backgrounds, aetiologies, and clinical pictures (Eysenck, 1970; Kendell, 1976; Poults and Bedford, 1976; Andreassen and Winokur, 1979). They also disagree about the boundaries between depressive disorders and anxiety disorders, schizophrenia, personality disorders, and normality (Roth et al., 1972; Kay et al., 1976; Kendell and Gourlay, 1970a and b; Tsuang et al., 1976). There is no consensus as to what constitutes adequate validation of a nosological class; a variety of independent validators have been proposed, including genetic factors, outcome, responses to treatment, and neurochemical or neurophysiological markers (Guze, 1978; Blashfield and Draguns, 1976; Kupfer et al., 1975; Reider and Gerston, 1978).

Multivariate statistical techniques have been applied to the problem of developing a classification of depressive disorders (Friedman et al., 1963; Raskin and Crook, 1976; Overall et al., 1966). More recent studies have used cluster analysis (Paykel, 1971; Fleiss, 1972). Most investigators now agree that cluster analysis is the most appropriate method for developing classification systems ex nihilo (Fleiss et al., 1971; Everit, 1972; Fleiss and Zubin, 1969; Maxwell, 1971; Straus et al., 1973; Blashfield, 1976). Factor analysis yields a dimensional clustering of symptoms but cluster analysis a categorical clustering of individuals, and therefore classifications generated by cluster analysis are like those used for clinical medicine.

These multivariate statistical studies of the classification of depressive disorders have led to surprisingly similar results. Paykel (1971) analyzed a group of 165 patients suffering from depression, including both in-patients and out-patients. He found an optimal solution at the level of four groups, which he named psychotic, anxious, hostile, and young depressives with personality disorder. In several subsequent studies (Paykel, 1972; Paykel et al., 1973) it was shown that these clusters were useful predictors of response to treatment, with the psychotic group responding best to amitriptyline. Using the less satisfactory method of inverse factor analysis Overall et al. (1966) identified three groups, which they called retarded, anxious, and hostile; the anxious group responded best to thioridazine, while the retarded group responded best to imipramine. Raskin and Crook (1976) identified four different groups which were called agitated, neurotic, endogenous, and personality-disordered depressives; however they were unable to identify any significant differences in response to treatment among the four groups.

While most clinical systems for classifying depression tend to be dichotomous and binary (endogenous vs. reactive, psychotic vs. neurotic, agitated vs. retarded), the studies using cluster analysis have consistently identified optimal solutions with three or four groups. All studies have identified a group which corresponds to endogenous depression but whether mathematical artefact or clinical reality remains in question for all groups. With the exception of examining response to treatment, the cluster analysis studies have not addressed the issue of finding independent validators such as course, prognosis, response to treatment, family history, or an independently derived classification system which employs diagnostic criteria.


---

Nancy C. Andreasen, M.D., Ph.D., Associate Professor of Psychiatry
William M. Grove, M.A., Research Associate
Ralph Maurer, M.D., Assistant Professor

Department of Psychiatry, University of Iowa, College of Medicine, 500 Newton Road, Iowa City, Iowa 52242, USA

* Reprint requests.

(Received 13 November 1979; revised 11 March 1980)
The present investigation uses the technique of cluster analysis to examine the classification of depression. Since all the patients were evaluated by a structured interview and diagnosed independently, criterion-based diagnoses can be compared with the mathematically derived classification system. Data concerning course and family history were also systematically collected.

**Method**

**Subjects:** All subjects in this study were evaluated in the pilot project of the NIMH Collaborative Studies of the Psychobiology of Depression in Boston, Iowa, New York, and St Louis. This study has been previously described (Katz et al., 1979), and the characteristics of its sample have been discussed in a previous publication (Andreasen and Winokur, 1979). Of the 150 patients in the original study, subjects were selected for this study who carried Research Diagnostic Criteria (RDC) (Spitzer et al., 1978) diagnoses of either definite major depressive disorder ($N = 78$) or definite minor depressive disorder ($N = 8$) for the current episode. All were in-patients. Mean age of the major depressives was 33.5 (SD = 15.0) while mean age for the minor depressives was 29.9 (SD = 10.1). Of the major depressives were 23 male and 55 female, while four minor depressives were male and four female. Mean educational attainment for major depressives was 12.8 years (SD = 2.8), and for minor depressives averaged 11.5 years (SD = 2.3).

**Measurements:** Interviewers used the Schedule for Affective Disorders and Schizophrenia (SADS) (Endicott et al., 1978). The current section of this interview consists of items concerning onset, course, and symptomatology of the episode and the past section provides information concerning previous illnesses. Seventy-one items were selected from the current section of the SADS to use for the cluster analysis. These items consisted of all cross-sectional symptoms contributing to a diagnosis of either major or minor depressive disorder. They included items relating to depressed mood and ideation, endogenous features, suicidal ideation and behaviour, anxiety symptoms, manic symptoms, and psychotic features; other symptoms included were obsessions and compulsions, alcohol and drug abuse, temper outbursts, antisocial behaviour, and disorientation. Longitudinal items covering onset and course were not included in the initial clustering procedures so that they could be used as potential independent validators for the categories. Subjects were also interviewed concerning family history using the Family History—Research Diagnostic Criteria (FH—RDC) (Andreasen et al., 1977) and information was collected concerning treatment received. These data were also not used for the clustering procedure, but were looked upon as potential independent validators.

**Statistical methods:** Since cluster analysis comprises a number of statistical techniques, the user must identify the method best suited to classify subjects from the population which he is studying. The major methodological problems in cluster analysis include choice of variables to generate the clusters, the type of clustering procedure (hierarchical vs. overlapping or partitioning, agglomerative vs. divisive), the mathematical criterion used to determine group membership and the appropriate number of clusters to generate. Deciding whether the clusters ultimately generated have any independent meaning or validity is perhaps the most important matter of all and it is the responsibility of the user to determine whether these clusters add to our knowledge concerning psychiatric classification.

Our decision to derive the clusters from the 71 SADS items covering current cross-sectional symptomatology was an arbitrary but heuristic strategy. This data base is the one commonly used for diagnostic purposes in clinical practice, although course and onset are often taken into account as well. Using only a cross-sectional approach imitates a clinical approach rather closely; it permits later use of longitudinal data concerning course, family history, and treatment to be used independently to test the predictive validity of the clusters.

A hierarchical agglomerative method, sometimes referred to as Ward’s method (Ward, 1963), was initially selected to develop the clusters using the CLUSTAN IC program (Wishart, 1975). In this programme the mathematical criterion for determining group
membership is to minimize the error sum of squares (the dissimilarity coefficient used), that is to assign patients to groups in such a way that the within-group sum of squares is as small as possible. This method tends to find tight clusters of approximately equal size of relatively homogeneous individuals. This was considered advantageous for several reasons. If a method generates clusters consisting of only three or four patients, these may occur so infrequently in a clinical population that they would have little meaning. Most of the current emphasis in psychiatric classification stresses the importance of identifying diagnostic classes which are as homogeneous as possible, particularly if the long-term goal is to identify biological markers.

Several investigators have pointed out that using a single method for clustering may produce clusters which are unstable. That is, patients will be classified one way by one method and another way by another method (Strauss et al., 1973; Evett, 1972). In order to avoid such instability, patients were reallocated from cluster to cluster by an iterative reassignment program which removes each person from his cluster and determines whether he would fit better in another. This reallocation was conducted three times from different starting clusters in order to ensure an optimal solution. These three reallocated solutions were compared by cross-tabulation. Sixty-nine of the 86 patients were allocated to the same cluster in each of the three solutions; the 69 patients who were stable across solutions probably represent the core members of their respective clusters and are probably most representative of their modal cluster types. For this reason, further analyses to identify cluster make-up and to evaluate predictive validity were conducted on those 69 patients.

A final difficulty in cluster analysis is to decide on the correct number of clusters to retain. This problem was resolved by examining agreement across the three iterative reallocations described above; up to ten cluster solutions were compared across the three allocations. Agreement between the three different solutions was examined as the number of clusters increased. There was a drop in agreement between a three-cluster and a four-cluster solution, indicating that a three-cluster solution was most stable. Therefore, the three-cluster solution was selected for further analysis.

**Results**

The SADS interview is divided into eight content-area scales: depressed mood and ideation, endogenous features, associated features, suicidal ideation, anxiety, manic syndrome, delusions and hallucinations, and formal thought disorder. In the initial data analysis, the scores of two groups were compared both on individual items and overall scale scores. Some results of these comparisons are summarized below.

Most of the differences between the three groups appeared in just areas tapped by the SADS evaluation: depressed mood and ideation, endogenous features, and suicidal ideation and behavior. The three groups differ significantly (P < .001) on three items grouped under depressed mood: overall severity of depressed mood, negative evaluation of self, and discouragement. In each instance, cluster 1 has the greatest severity and cluster 3 the least severity of symptoms. And, in each instance, most of the difference is between clusters 1 and 2 vs. cluster 3.

A theoretically relevant distinction between groups is one based on endogenous symptoms. Table 1 shows the differences between the three groups on endogenous features. The three groups differ significantly on approximately eight items (depending on the cut-off level used to determine statistical significance). These include: terminal insomnia, loss of appetite, weight loss, decreased interest, psychomotor retardation, nonreactivity of mood with changes in circumstances, anxiety to experience pleasure, and diurnal variation with symptoms worse in the morning. Cluster 1 nearly always has the most severe symptoms and cluster 3 the least severe. In five out of eight instances, the difference is primarily between cluster 1 vs. 2 and 3. These five instances involve the symptoms often used to diagnose endogenous depression: terminal insomnia, loss of appetite, weight loss, psychomotor retardation, and diurnal variation. The other three differences, for which clusters 1 and 2 are similar to one another but different from cluster 3, involve: including decreased mood, and inability to...
Cluster solution was most rare-cluster solution was sis.

**Table 1**

Symptoms from SADS endogenous feature scale by cluster core group membership

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Cluster 1 (N = 32)</th>
<th>Cluster 2 (N = 19)</th>
<th>Cluster 3 (N = 18)</th>
<th>F**</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of mood</td>
<td>3.03 (1.45)</td>
<td>2.58 (1.54)</td>
<td>2.33 (1.57)</td>
<td>1.36</td>
<td>.&lt;.1</td>
</tr>
<tr>
<td>Lack of specific concerns</td>
<td>4.00 (1.76)</td>
<td>2.95 (1.47)</td>
<td>2.89 (1.84)</td>
<td>3.44</td>
<td>.048</td>
</tr>
<tr>
<td>Blames self</td>
<td>3.13 (1.41)</td>
<td>2.79 (1.08)</td>
<td>2.61 (1.09)</td>
<td>1.08</td>
<td>.&lt;.1</td>
</tr>
<tr>
<td>Middle insomnia</td>
<td>3.16 (1.90)</td>
<td>2.32 (1.83)</td>
<td>2.33 (1.75)</td>
<td>1.75</td>
<td>.&lt;.1</td>
</tr>
<tr>
<td>Terminal insomnia</td>
<td>3.91 (2.15)</td>
<td>2.16 (1.89)</td>
<td>1.83 (1.76)</td>
<td>8.05</td>
<td>.&lt;.001</td>
</tr>
<tr>
<td>Loss of appetite</td>
<td>4.51 (1.38)</td>
<td>2.21 (1.51)</td>
<td>3.06 (1.51)</td>
<td>13.22</td>
<td>.&lt;.001</td>
</tr>
<tr>
<td>Weight loss</td>
<td>3.22 (1.60)</td>
<td>2.16 (1.50)</td>
<td>2.33 (1.85)</td>
<td>3.08</td>
<td>.050</td>
</tr>
<tr>
<td>Decreased interest</td>
<td>5.16 (0.99)</td>
<td>4.79 (1.18)</td>
<td>2.72 (1.81)</td>
<td>21.36</td>
<td>.&lt;.001</td>
</tr>
<tr>
<td>Agitation</td>
<td>2.41 (1.43)</td>
<td>1.79 (0.98)</td>
<td>2.89 (1.29)</td>
<td>1.55</td>
<td>.&lt;.1</td>
</tr>
<tr>
<td>Retardation</td>
<td>3.09 (1.30)</td>
<td>1.95 (1.03)</td>
<td>1.55 (0.92)</td>
<td>12.30</td>
<td>.&lt;.001</td>
</tr>
<tr>
<td>Reactivity</td>
<td>4.44 (1.52)</td>
<td>3.74 (1.59)</td>
<td>2.33 (1.57)</td>
<td>10.88</td>
<td>.&lt;.001</td>
</tr>
<tr>
<td>Pleasure capacity</td>
<td>4.31 (1.89)</td>
<td>4.53 (1.50)</td>
<td>2.22 (1.80)</td>
<td>10.09</td>
<td>.&lt;.001</td>
</tr>
<tr>
<td>Worse in morning</td>
<td>2.22 (1.36)</td>
<td>1.16 (0.69)</td>
<td>1.44 (1.04)</td>
<td>5.99</td>
<td>.003</td>
</tr>
</tbody>
</table>

*Since these symptoms were used to form the clusters, the clusters differ on them a priori. The 'F' statistic therefore lacks in F distribution, and P values are printed only for illustrative purposes.

From cluster 3, involve psychological symptoms, including decreased interest, nonreactivity of mood, and inability to experience pleasure. Taken together with differences on depressed mood items, these results suggest that the severity of the illness accounts for much of the difference between the three groups but that cluster 1 may correspond to the concept of endogenous depression.

Other item analyses indicate that cluster 2 is highest on symptoms which suggest emotional lability and reactivity, including subjective anger, temper outbursts, and alcohol abuse during the episodes of depression. In these three instances, cluster 2 is distinct from 1 and 3 (P < .001). On the other hand, on two cognitive symptoms, indecisiveness and difficulty in concentrating, we find the familiar pattern of severity with cluster 1 being most severe and cluster 3 being least severe (P < .001). On the former two of these cognitive symptoms, cluster 1 differs from 2 and 3; on psychic anxiety, cluster 1 and 2 are similar and differ from 3. These data suggest that, while severity of symptoms may be a major difference between the three groups, they may also differ in a qualitative way. Cluster 2 appears to correspond in some ways to the concept of reactive or neurotic depression.

Together with depressed mood and endogenous features, the SADS scale on suicide ideation and behaviour is most successful in distinguishing between the three clusters. On this scale cluster 2 is consistently higher than clusters 1 and 3. On two items, suicidal tendencies and number of attempts, the major difference is cluster 2 vs. 1 and 3. On the other

![SADS Scale Score Profile by Cluster](image-url)

**Fig. 1.**
hand, cluster 2 resembles cluster 1 and differs from 3 in terms of seriousness of intent and danger to life. This suggests that both clusters 1 and 2 represent serious suicide risks, but the greater number of attempts by cluster 2 may be another indicator of the reactivity and volatility of this group.

Fig 1 portrays the scale scores of the three clusters on the eight SADS subscales. This may be viewed as a profile of the symptomatology of each of the three clusters. Scale scores have been standardized to have a mean of 50 and a standard deviation of 10 in the entire study sample in order to facilitate comparisons between scales.

These data suggest that the three clusters display distinct differences in severity of symptomatology and types of symptoms. Cluster 1 would appear to have more endogenous features and vegetative symptoms of depression than are accounted for by its overall severity of depression, while cluster 2 has distinctly more suicidal ideation and behaviour than one would expect, based on its apparent level of depression. While Fig 1 is enlightening, a simple visual inspection of the profiles does not allow a quantitative assessment of the relative contributions of severity and type of psychopathology to differences between clusters. Therefore, a multivariate analysis of variance on the symptom profiles was used to examine the degree to which inter-cluster differences depend on overall severity of psychopathology vs. symptom profile differences.

The first discriminant function is the largest multivariate component of between-group differences. A graph of the scores on the first discriminant function between the clusters is given in Fig 2. A number of investigators have noted that, if groups represent true subtypes, the distribution of this discriminant function should be bimodal (or trimodal in the case of three groups). As Kendall (1966) has said, if there are two distinct types of depression, the blacks and whites should be more common than the grays. When this distribution has been examined in the past, only Carney et al. (1965) have been able to obtain a bimodal distribution suggesting of separate and distinct subtypes. The distribution of the discriminant scores best separating the three clusters, as given in Fig 2, is clearly unimodal. Although Fleiss (1972) has shown that such distributions cannot be interpreted simply by eye, the distribution does seem to suggest that the three groups represent a continuum.

However, a less simple interpretation of the data is suggested by the results of the multivariate analysis of variance, given in Table IV. A test of parallelism of symptom profiles (i.e., all differences are ones of elevation) is given by a multivariate analysis of variance conducted on the slope measures between adjacent SADS subscales. For example, in Fig 1, where it appears that clusters 1 and 2 have a different pattern on vegetative symptoms of depression and suicidal ideation and behaviour, this profile difference is quantified by the difference in slopes between the lines extending from Scale 3 to Scale 4 in each of the clusters. The differences between symptom profile shapes are highly significant ($P < .001$). Severity of psychopathology, independent of profile shape, is quantified by calculating for each patient the average of his or her SADS scale scores. A patient with high profile elevation may have one very high scale or several moderately high scales. The difference between clusters on profile elevation is also highly significant ($P < .001$).

Given that clusters differ on profile elevation and shape measures, one may ask whether severity of psychopathology than its type in diagnosis. An answer is that severe accounts for 62 per cent of the total variance (profile shape between-cluster dispersion). Diagnoses were made using the SADS interview and Diagnostic Criteria (ICD-10). The clusters provide additional information on clinical make-up of the potential independent variables.

**RDC subtypes**

- **Major depression**
- **Minor depression**
- **Primary (P+D)**
- **Secondary (P+D)**
- **Recurrent (P+D)**
- **Psychotic (P+D)**
- **Incapacitating (P+D)**
- **Endogenous (P+D)**
- **Endogenous (P)**
- **Endogenous (D)**
- **Retarded (P+D)**
- **Agitated (P+D)**
- **Situational (P+D)**
- **Simple (P+D)**

* Cramer's V statistic indicates that the data is not complete
** $P$ = Probable; $D$ = Definite

**Treatment**

- **ECT**
- **Antidepressant**
- **Lithium**
- **Minor tranquilizer**
- **Major tranquilizer**
- **Other drug**

* Numbers total to more than 100
severity of psychopathology is more important than its type in distinguishing clusters. The answer is that severity (profile elevation) accounts for 62 per cent and type of symptomatology (profile shape) for 38 per cent of the between-cluster dispersion.

Diagnoses were made on all subjects at the time of the SADS interview using the Research Diagnostic Criteria (RDC). The relationship between the clusters and RDC diagnoses provides additional information concerning the clinical make-up of the three groups and is a potential independent validator. These data are portrayed in Table II. They further bear out the characterization of cluster 1 as endogenous and cluster 2 as reactive or neurotic. The three clusters differ significantly on the number of patients diagnosed as either definite or probably endogenous on the RDC. Both cluster 2 and cluster 3 contain no definite endogenous depressions, while 59 per cent of cluster 1 are definitely endogenous. Cluster 1 also contains more retarded depressions than 2 or 3. On the other hand, cluster 2 contains significantly more situational depressions. Finally, cluster 3 contains significantly more minor depressions than clusters 1 and 2.

**Table II**

<table>
<thead>
<tr>
<th>RDC subtypes**</th>
<th>Cluster 1 (N = 32)</th>
<th>%</th>
<th>Cluster 2 (N = 19)</th>
<th>%</th>
<th>Cluster 3 (N = 18)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major depression</td>
<td>32</td>
<td>100</td>
<td>17</td>
<td>89</td>
<td>12</td>
<td>67</td>
</tr>
<tr>
<td>Minor depression</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>6</td>
<td>33</td>
</tr>
<tr>
<td>Primary (P+D)</td>
<td>22</td>
<td>69</td>
<td>8</td>
<td>42</td>
<td>10</td>
<td>56</td>
</tr>
<tr>
<td>Secondary (P+D)</td>
<td>10</td>
<td>31</td>
<td>9</td>
<td>47</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Recurrent (P+D)</td>
<td>17</td>
<td>53</td>
<td>9</td>
<td>47</td>
<td>8</td>
<td>44</td>
</tr>
<tr>
<td>Psychotic (P+D)</td>
<td>9</td>
<td>28</td>
<td>4</td>
<td>21</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>Incapacitating (P+D)</td>
<td>21</td>
<td>66</td>
<td>6</td>
<td>32</td>
<td>6</td>
<td>33</td>
</tr>
<tr>
<td>Endogenous (P)</td>
<td>27</td>
<td>84</td>
<td>10</td>
<td>53</td>
<td>7</td>
<td>39</td>
</tr>
<tr>
<td>Endogenous (D)</td>
<td>8</td>
<td>25</td>
<td>10</td>
<td>53</td>
<td>7</td>
<td>39</td>
</tr>
<tr>
<td>Retarded (P+D)</td>
<td>19</td>
<td>59</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Agitated (P+D)</td>
<td>14</td>
<td>44</td>
<td>3</td>
<td>16</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Situational (P+D)</td>
<td>7</td>
<td>22</td>
<td>2</td>
<td>11</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>Simple (P+D)</td>
<td>20</td>
<td>63</td>
<td>12</td>
<td>63</td>
<td>9</td>
<td>50</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 12.52, P = .002, V = .426 \]

* Cramér’s V statistic of agreement. It ranges from 0 (no association) to 1 (all cells are either completely empty or completely full). ** P = Probable; D = Definite.

**Table III**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Cluster 1 (N = 32)</th>
<th>%</th>
<th>Cluster 2 (N = 19)</th>
<th>%</th>
<th>Cluster 3 (N = 18)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECT</td>
<td>4</td>
<td>13</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>Antidepressant</td>
<td>26</td>
<td>81</td>
<td>8</td>
<td>42</td>
<td>12</td>
<td>67</td>
</tr>
<tr>
<td>Lithium</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Minor tranquilizer</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>37</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>Major tranquilizer</td>
<td>9</td>
<td>28</td>
<td>6</td>
<td>32</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Other drug</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 3.97, P = .050 \]

* Numbers total to more than cluster Ns due to patients receiving multiple treatments.
CLUSTER ANALYSIS AND THE CLASSIFICATION OF DEPRESSION

The three clusters were also compared in terms of course, family history, and treatment received. The three groups did not differ on duration of the episode at time of evaluation, rapidity of onset, and rater's judgment of life stress just before onset. The data concerning duration and onset were rated on a nine-point scale rather than in terms of actual weeks or months, and this may have caused differences to be less pronounced. Because a comprehensive follow-up study was not done, adequate data are not available on outcome, which is perhaps the most important course variable. The data available tend to be in the directions expected, given the symptomatologic nature of the clusters. Clusters 1 and 2 tend to have no more rapid onset and cluster 2 a longer duration. Cluster 2 also tends to have an onset associated with stressful life events, with such events occurring in 95 per cent of cases (vs. 69 per cent in cluster 1 and 72 per cent in cluster 3).

Family history was also considered as a potential validating variable but no significant differences were noted between clusters 1 and 2 in this area. Clusters 1 and 2 had high familial prevalence of affective disorder, with 16 per cent of relatives in each group having a positive history vs. 7 per cent in cluster 3. This difference is significant (P < .025). Since these data are based primarily on family history data, a method which tends to yield a high rate of false negatives, they should only be considered tentative.

The three clusters were also compared in terms of treatment received. These data are summarized in Table III. Treatment received is some indication of the way in which clinicians perceived the illness which these patients had. As Table III indicates, patients in cluster 1 tended to be treated primarily with antidepressants, while patients in cluster 2 were more often treated with minor tranquilizers. It is interesting to note, however, that cluster 3 tended to receive relatively more of all treatments, including antidepressants and even ECT. This is rather surprising in view of the fact that cluster 3 consists of patients with the mildest symptoms. One can only speculate that these patients were treated with a number of different somatic therapies because clinicians found their management difficult.

Discussion

In some respects, these data are surprisingly similar to the results of other studies which have used cluster analysis as an aid to classification. These studies have consistently identified three or four groups, which have always included a group which could be characterized as endogenous or retarded. Paykel (1971), Lorr et al.

### Table IV

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sums of squares between</th>
<th>Sums of squares within</th>
<th>Univariate F (2,66)</th>
<th>Characteristic roots of ((W^{-1}B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile elevation (average of all 8 standardized SADS scale scores)</td>
<td>2119.60</td>
<td>1515.19</td>
<td>46.16</td>
<td>1.40</td>
</tr>
<tr>
<td>Profile slope (2-1)</td>
<td>884.77</td>
<td>8458.45</td>
<td>3.44</td>
<td></td>
</tr>
<tr>
<td>Profile slope (3-2)</td>
<td>203.94</td>
<td>1998.61</td>
<td>3.37</td>
<td></td>
</tr>
<tr>
<td>Profile slope (4-3)</td>
<td>3181.84</td>
<td>6201.97</td>
<td>16.12</td>
<td></td>
</tr>
<tr>
<td>Profile slope (5-4)</td>
<td>752.23</td>
<td>11511.95</td>
<td>2.23</td>
<td>1.77</td>
</tr>
<tr>
<td>Profile slope (6-5)</td>
<td>250.29</td>
<td>12122.42</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Profile slope (7-6)</td>
<td>97.37</td>
<td>13096.14</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Profile slope (8-7)</td>
<td>580.17</td>
<td>13814.47</td>
<td>1.41</td>
<td></td>
</tr>
</tbody>
</table>

* Since all symptoms defining these scales were used to form the clusters, the clusters differ on the scales *a priori*. The characteristic root statistics therefore lack their theoretical distribution, and P values are given only for illustrative purposes. P was derived from the distribution of trace \((B (B+W)^{-1}) = 0.989\) for the profile slopes.
yield a high rate of
uld only be considered
were also compared in
eived. These data are
II. Treatment received
way in which clinicians
ich these patients had.
patients in cluster 1
primarily with an
tents in cluster 2 were
or minor tranquilizers.
however, that cluster 3
ively more of all treat-
pressants and even ECT.
in view of the fact that
ients with the mildest
t to speculate that these
th a number of different
use clinicians found their
(1973), and Overall et al (1965)
identify additional groups which are characterized as
anxious or hostile, while Fleiss (1972) and
Raskin and Crook (1976) have identified
additional groups which they characterize as
neurotic, agitated, and personality-disordered.
This study clearly replicates the finding of an
endogenous group, while cluster 2 appears to
Correspond roughly to a neurotic or reactive
which is probably a combination of
Paykel’s and Overall’s hostile and anxious
groups. Cluster 3 appears to consist of milder
depressions which may correspond to Paykel’s
cluster 4 and to Raskin and Crook’s cluster 4,
patients with personality disorder, and to the
categories of chronic minor depression or
lative personality. Thus, the present study
seems to provide further support for the concept
or endogenous depression, which has clearly
emerged from all studies.

This study provides an independent confirm-
ation of the existence of several RDC sub-
types of depressive disorder. In particular,
major depression, minor depression, endogenous
depression, and situational depression have in a
sense been independently identified by the three
clusters. Further, the differences in assigned
treatment also suggest that clinicians perceived
the groups somewhat differently.

A number of reservations should be men-
tioned. First, the three groups did not differ
significantly in terms of family history or course
variables, which are often considered important
independent validators of a classification. This
may be due to basic problems in the data base,
such as lack of information concerning long-
term prognosis or lack of complete family
study data. Second, there is some circularity in
the method, in that the variables used to
identify the clusters are also those used to
define the subtypes within the RDC clas-
sification system. Third, the similarity of the
findings to those of other cluster analysis
studies may be more apparent than real, in that
the similarity resides in the comparison of
superficial descriptions. The method should be
repeated on a subsequent sample, larger in size
and with more complete information concerning
course, prognosis, and response to treatment, in
order to determine whether the clusters are
truly stable and adequately validated.

One may question whether these clusters
represent true subtypes at all. Since it is
inherent in cluster analysis that clusters will be
produced, their existence does not prove that
they represent true subtypes; it could be argued
that the major differences between the three
clusters are quantitative rather than qualitative,
and that the three clusters simply correspond to
mild, moderate, and severe forms of depression.
However, the results of the analysis of profile
elevation (severity) and profile shape (the
qualitative differences in psychopathology)
establish that both differences in severity and in
symptom patterns distinguish the clusters.

In summary, the results of this investigation
suggest that two different components are
important in the classification of depression:
severity and type of symptoms. The difference
between subtypes is probably both qualitative
and quantitative, both continuous and cate-
gorical. Although this conclusion is complex
and paradoxical, it does reflect the reality of
clinical practice in which endogenous depression
is the category which enjoys widest acceptance
as a subtype likely to be predictive of outcome
and response to treatment. In this study,
endogenous depression is the subtype which
most clearly emerges as a separate cluster. On
the other hand, attempts to identify and define
other subtypes of depression have been the
subject of controversy in both clinical practice
and in research. In this study, the two non-
edogenous clusters do not clearly correspond
to accepted clinical or research categories, are
not validated by use of specific treatments, and
do not emerge as robustly as independent
categories. This study suggests that further
examination of non-endogenous depression is
needed, but that subtypes within this category
may be quite difficult to define independently
from a severity dimension.

Acknowledgements

This work was done as part of the NIMH-Clinical
Research Branch Collaborative Program on the Psycho-
biology of Depression: Clinical Studies, and was completed
with the cooperation and participation of the Col-
laborative Program investigators: Martin M. Katz, Ph.D.,
and Gerald L. Klerman, M.D., co-chairmen; Robert
References


