Brit. J. Psychiat. (1980), **137**, 256–265

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; Foulds and Bedford, 1976; Andreasen 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 Gershon, 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; Everitt, 1972; Fleiss and Zubin, 1969; Maxwell, 1971; Strauss *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.

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

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 methodologic 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 1C program (Wishart, 1975). In this programme the mathematical criterion for determining group

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

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; Everitt, 1972). In order to avoid such instability, patients were reallocated from cluster to cluster by an iterative reallocation programme 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 three 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 three areas tapped by the SADS evaluation: depressed mood and ideation, endogenous features, and suicidal ideation and behaviour. 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 lowest 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 I 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, inability 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

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

TABLE I

Symptoms from SADS endogenous feature scale by cluster core group membership

Symptom	Cluster 1 (N = 32)		Cluster 2 (N = 19)		Cluster 3 (N = 18)		'F'*	P	Homogeneous subsets
	x̄	SD	x̄	SD	x̄	SD			
Quality of mood	3.03	(1.45)	2.58	(1.54)	2.33	(1.57)	1.36	>.10	
Lack of specific concerns	4.00	(1.76)	2.95	(1.47)	2.89	(1.84)	3.44	.048	1>2, 3
Blames self	3.13	(1.41)	2.79	(1.08)	2.61	(1.09)	1.08	>.10	
Middle insomnia	3.16	(1.90)	2.32	(1.83)	2.33	(1.75)	1.75	>.10	
Terminal insomnia	3.91	(2.15)	2.16	(1.89)	1.83	(1.76)	8.05	<.001	1>2, 3
Loss of appetite	4.31	(1.38)	2.21	(1.51)	3.06	(1.51)	13.22	<.001	1>3, 2
Weight loss	3.22	(1.60)	2.16	(1.50)	2.33	(1.85)	3.08	.050	
Decreased interest	5.16	(0.99)	4.79	(1.18)	2.72	(1.81)	21.36	<.001	1, 2>3
Agitation	2.41	(1.43)	1.79	(0.98)	2.39	(1.29)	1.55	>.10	
Retardation	3.09	(1.30)	1.95	(1.03)	1.55	(0.92)	12.30	<.001	1>2, 3
Reactivity	4.44	(1.52)	3.74	(1.59)	2.33	(1.57)	10.88	<.001	1, 2>3
Pleasure capacity	4.31	(1.89)	4.53	(1.50)	2.22	(1.80)	10.09	<.001	2, 1>3
Worse in morning	2.22	(1.36)	1.16	(0.69)	1.44	(1.04)	5.99	.003	1>3, 2

* 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 experiecne 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

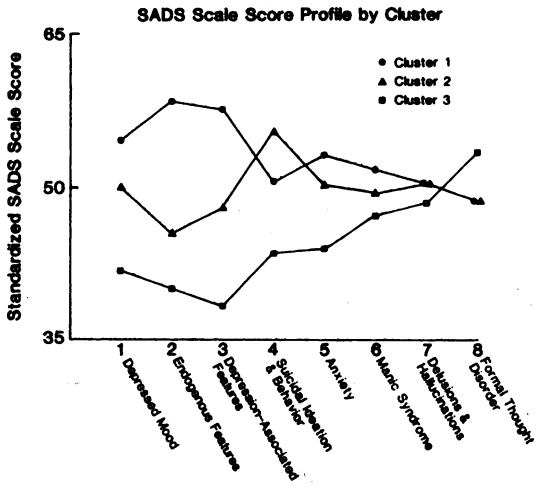


FIG 1.

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

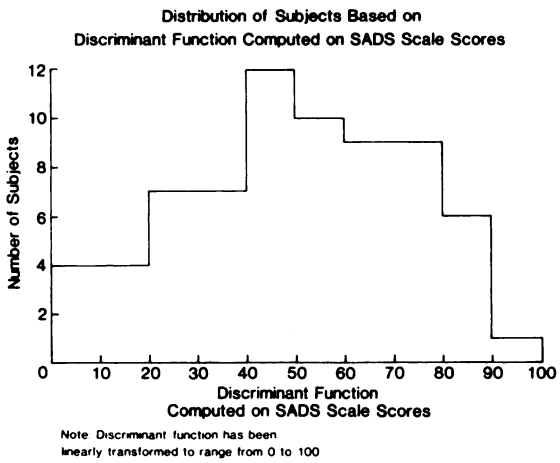


FIG 2.

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 Kendell (1968) 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 suggestive 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

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

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

Agreement of subtypes of depression by research diagnostic criteria, with cluter core group membership

	Number of members with subtypes								
	Cluster 1		Cluster 2		Cluster 3				
RDC subtypes**	(N = 32)	%	(N = 19)	%	(N = 18)	%	χ^2	P	V*
Major depression	32	100	17	89	12	67	12.52	.002	.426
Minor depression	0	0	2	11	6	33			
Primary (P+D)	22	69	8	42	10	56	4.40	>.10	.253
Secondary (P+D)	10	31	9	47	2	11			
Recurrent (P+D)	17	53	9	47	8	44	0.72	>.10	.102
Psychotic (P+D)	9	28	4	21	3	17	0.13	>.10	.043
Incapacitating (P+D)	21	66	6	32	6	33	4.13	>.10	.245
Endogenous (P+D)	27	84	10	53	7	39	5.02	.081	.270
Endogenous (P)	8	25	10	53	7	39	7.11	.029	.321
Endogenous (D)	19	59	0	0	0	0	25.01	<.001	.602
Retarded (P+D)	14	44	3	16	2	11	4.99	.082	.269
Agitated (P+D)	7	22	2	11	5	28	3.60	>.10	.228
Situational (P+D)	12	38	13	68	4	22	7.97	.019	.340
Simple (P+D)	20	63	12	63	9	50	0.74	>.10	.104

* Cramer'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

*Treatments received by cluster members**

	Cluster 1		Cluster 2		Cluster 3			
Treatment	(N = 32)	%	(N = 19)	%	(N = 18)	%	χ^2	P
ECT	4	13	1	5	5	28	3.97	>.10
Antidepressant	26	81	8	42	12	67	8.22	.016
Lithium	3	9	0	0	4	22	5.05	.080
Minor tranquillizer	3	9	7	37	5	28	5.81	.055
Major tranquillizer	9	28	6	32	4	22	0.42	>.10
Other drug	2	6	0	0	1	6	3.00	>.10

* Numbers total to more than cluster Ns due to patients receiving multiple treatments.

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

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 a 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 tranquillizers. 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

Results of multivariate analysis of variance for differences between clusters on SADS scales

Variable	Sums of squares between	Sums of squares within	Univariate F (2,66)	Characteristic roots of ($W^{-1}B$) 1	2	P*
Profile elevation (average of all 8 standardized SADS scale scores)	2119.60	1515.19	46.16	1.40	0	<.001
Profile slope (2–1)	884.77	8485.45	3.44			
Profile slope (3–2)	203.94	1998.61	3.37			
Profile slope (4–3)	3181.84	6512.97	16.12			
Profile slope (5–4)	752.23	11151.93	2.23	1.77	0.25	<.001
Profile slope (6–5)	250.29	12122.42	0.68			
Profile slope (7–6)	97.37	13096.14	0.25			
Profile slope (8–7)	588.17	13814.47	1.41			

* 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.

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

(1973), and Overall *et al* (1966) have identified 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 group 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 concepts of chronic minor depression or depressive 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 confirmation of the existence of several RDC subtypes 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 mentioned. 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 classification 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 groups 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 categorical. 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-endogenous 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 Psychobiology of Depression: Clinical Studies, and was completed with the cooperation and participation of the Collaborative Program investigators: Martin M. Katz, Ph.D., and Gerald L. Klerman, M.D., co-chairmen; Robert

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

M. A. Hirschfeld, M.D., Coordinator; Martin B. Keller, M.D., Robert W. Shapiro, M.D., and James Barrett, M.D. (Boston); Jan A. Fawcett, M.D., and William A. Scheftner, M.D. (Chicago); George Winokur, M.D., and Nancy C. Andreasen, M.D. (Iowa City); Robert L. Spitzer, M.D., and Jean Endicott, Ph.D. (New York); and Eli Robins, M.D., Paula J. Clayton, M.D., Theodore Reich, M.D., Amos Welner, M.D., and Jack Croghan, M.D. (St Louis). Useful comments on this manuscript were also provided by Roger Blashfield, Will Carpenter, Joseph Fleiss, and Peter A. Lachenbruch. Funded in part by NIMH Grant MH225416-01-06.

References

ANDREASEN, N. C., ENDICOTT, J., SPITZER, R. L. & WINOKUR, G. (1977) The family history method using diagnostic criteria: Reliability and validity. *Archives of General Psychiatry*, **34**, 1229–35.

—— & WINOKUR, G. (1979) Newer experimental methods for classifying depression. *Archives of General Psychiatry*, **36**, 447–52.

BLASHFIELD, R. K. (1976) Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, **83**, 377–88.

—— & MOREY, L. C. (1979) The classification of depression through cluster analysis. *Comprehensive Psychiatry*, **20**, 516–27.

—— & DRAGUNS, J. G. (1976) Evaluative criteria for psychiatric classification. *Journal of Abnormal Psychology*, **85**, 140–50.

CARNEY, M. W. P., ROTH, M. & GARSIDE, R. F. (1965) The diagnosis of depressive syndromes and the prediction of ECT response. *British Journal of Psychiatry*, **111**, 659–74.

ENDICOTT, J. & SPITZER, R. L. (1978) A diagnostic interview: The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, **35**, 837–44.

EVERITT, B. S. (1972) Cluster analysis: A brief discussion of some of the problems. *British Journal of Psychiatry*, **120**, 143–5.

EYSENCK, H. J. (1970) The classification of depressive illness. *British Journal of Psychiatry*, **117**, 241–50.

FLEISS, J. L. (1972) Classification of the depressive disorders by numerical typology. *Journal of Psychiatric Research*, **9**, 141–53.

—— & ZUBIN, J. (1969) On the methods and theory of clustering. *Multivariate Behavioral Research*, **4**, 235–50.

—— LAWLOR, W., PLATMAN, S. R. & FIEVE, R. R. (1971) On the use of inverted factor analysis for generating typologies. *Journal of Abnormal Psychology*, **77**, 127–32.

FOULDS, G. A. & BEDFORD, A. (1976) Classification of depressive illness: A re-evaluation. *Psychological Medicine*, **6**, 15–19.

FRIEDMAN, A. S., COWTIZ, B., COHEN, H. W. & GRANICK, S. (1963) Syndromes and themes of psychotic depression: A factor analysis. *Archives of General Psychiatry*, **9**, 504–9.

GUZE, S. B. (1978) Validating criteria for psychiatric diagnosis: The Washington University Approach. In *Psychiatric Diagnosis: Exploration of Biological Predictors* (eds. H. S. Akiskal and W. L. Webb). New York and London: Spectrum Publications.

KATZ, M. M., SECUNDA, S. K., HIRSCHFIELD, R. M. A. & KOSLOW, S. H. (1979) NIMH Clinical Research Branch Collaborative Program on the Psychobiology of Depression. *Archives of General Psychiatry*, **36**, 765–71.

KAY, D. W. K., COOPER, A. F., GARSIDE, R. F. & ROTH, M. (1976) The differentiation of paranoid from affective psychoses by patients' premorbid characteristics. *British Journal of Psychiatry*, **129**, 207–15.

KENDELL, R. E. (1968) *The Classification of Depressive Illness*. London: Oxford University Press.

—— (1976) The classification of depressions: A review of contemporary confusion. *British Journal of Psychiatry*, **129**, 15–28.

—— & GOURLAY, J. (1970) The clinical distinction between the affective psychoses and schizophrenia. *British Journal of Psychiatry*, **117**, 261–6.

—— —— (1970) The clinical distinction between psychotic and neurotic depressions. *British Journal of Psychiatry*, **117**, 257–66.

KUPFER, D. J., PICKAR, D., HIMMELHOCH, J. M. & DETRE, T. P. (1975) Are there two types of unipolar depression? *Archives of General Psychiatry*, **32**, 866–71.

LORR, M., POKORNY, A. D. & KLETT, C. J. (1973) Three depressive types. *Journal of Clinical Psychology*, **29**, 290–4.

MAXWELL, A. E. (1971) Multivariate statistical methods and classification problems. *British Journal of Psychiatry*, **119**, 121–7.

OVERALL, J. E., HOLLISTER, L. E., JOHNSON, M. & PENNINGTON, V. (1966) Nosology of depression and differential response to drugs. *Journal of the American Medical Association*, **195**, 946–50.

PAYKEL, E. S. (1972) Correlates of a depressive typology. *Archives of General Psychiatry*, **27**, 203–10.

—— (1971) Classification of depressed patients: A cluster analysis derived grouping. *British Journal of Psychiatry*, **118**, 257–88.

—— PRUSOFF, B. A., KLERMAN, G. L., HASKELL, D. & DIMASCIO, A. (1973) Clinical response to amitriptyline among depressed women. *The Journal of Nervous and Mental Disease*, **156**, 149–65.

RASKIN, A. & CROOK, T. H. (1976) The endogenous-neurotic distinction as a predictor of response to antidepressant drugs. *Psychological Medicine*, **6**, 59–70.

REIDER, R. O. & GERSHON, E. S. (1978) Genetic strategies in biological psychiatry. *Archives of General Psychiatry*, **35**, 866–73.

SPITZER, R. L., ENDICOTT, J. & ROBINS, E. (1978) Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry*, **35**, 773–82.

STRAUSS, J. S., BARTKO, J. J. & CARPENTER, W. T. JR. (1973) The use of clustering techniques for the classification of psychiatric patients. *British Journal of Psychiatry*, **122**, 531–40.

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press

TSUANG, M., DEMPSEY, M. & RAUSCHER, F. (1976) A study of atypical schizophrenia. *Archives of General Psychiatry*, **33**, 1157–60.

WARD, J. JR. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**, 236–44.

WISHART, D. (1975) *CLUSTAN IC User Manual* (*Release 1*). London: University College Computer Centre.

* 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*)

https://doi.org/10.1192/bjp.137.3.256 Published online by Cambridge University Press