Brit. J. Psychiat. (1971), 118, 275–88

Classification of Depressed Patients: A Cluster Analysis Derived Grouping

By E. S. PAYKEL

In recent years there has been a considerable interest in the classification of depressed patients, and in the use of multivariate statistical techniques as tools for classifying, with the hope of either validating traditional classifications or suggesting useful new ones. One technique employed for this purpose has been factor analysis. A number of studies have used this method to explore the overlapping endogenous-reactive or psychotic-neurotic dichotomies. Most have demonstrated a factor or factors reflecting such a dimension which appears to form a consistent, systematic source of variance in ratings of depressed patients (1, 5, 11, 15, 16, 18, 26, 31, 33).

A limitation of these studies is that most have been concerned with simple classifications of patients into two groups. Such dichotomies, although useful, do not appear to do full justice to the manifest heterogeneity of depressed patients which must be systematized if a unitary hypothesis is to be abandoned. Few studies have seriously explored the possibility of more than two groups. Kendell (16) found no evidence to support a separate group of involutional melancholics. Two American studies suggested multiple groupings. Grinker and his colleagues (8), in an extensive and detailed study of depressed patients, suggested four patterns of patients. Overall, *et al* (24) derived three groups, retarded, anxious and hostile depressives. These groups were subsequently found to have predictive value in treatment since they showed different responses to tricyclic antidepressants and phenothiazines (13).

Another problem, and one which may in part explain the concentration on two groups, lies in the extensive use of factor analysis of rating measures as the main statistical tool to explore the classification of individuals. This technique isolates dimensions among rating variables. These do not necessarily coincide with groups of patients. Moreover, these multiple dimensions are usually orthogonal so that variation along each is independent. Description of position on one such dimension says nothing as to position on any of the others, and has very different implications from membership of one of a set of mutually exclusive groups. The awkwardness of making such interpretations becomes particularly apparent when multiple factors are used to argue for multiple groups.

More recently, multivariate cluster analysis techniques have been described which are designed particularly to form groups of individuals (19, 36). In fact, the two studies referred to above which derived multiple groupings used variants of one such approach—factor analysis on correlations of patient profiles (Q correlations) rather than of rating items. Another recent study used a technique based on information theory and applied it to a mixed sample of psychiatric patients including depressives; the latter appeared to divide into endogenous and reactive groups (27).

This paper will report the results of applying another promising cluster analysis technique to a heterogeneous sample of 165 depressed patients from a variety of treatment settings. A previous factor analysis of these data revealed a relatively weak factor corresponding to the psychotic-neurotic dichtotomy (26); several other factors could not easily be interpreted. The results of the cluster analysis to be reported suggest a more extensive and readily interpretable classification into four groups, overlapping both with the psychotic-neurotic dichotomy and with other groupings described in the literature.

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

METHODS

A. *Subjects*

The subjects were consecutive patients with depression admitted to a number of different psychiatric units serving one urban area in the U.S.A. They included 100 out-patients, 30 day hospital patients, and 25 patients admitted to a brief emergency admission unit, all from a community mental health centre, together with a total of 65 in-patients admitted variously to the mental health centre, the psychiatric ward of a general hospital, a state mental hospital and a Veterans' Administration hospital. They were obtained by screening all admissions to these units. Full details of sampling and screening procedures have been described elsewhere (25). For the analyses to be presented here the total sample of 220 depressed patients was reduced to 165 by exclusion of one-quarter, stratified by treatment unit, but otherwise random.

The criteria for depression specified presence of psychiatric disorder in which the central feature was abnormal, persistent depressed affect characterized by feelings of depression, sadness, or a tendency to cry, which might be accompanied by guilt, worthlessness, hopelessness, suicidal feelings, or appearance of depression at interview. The illness was of at least one week's duration, and sufficiently severe for the overall illness to be rated 2 (mild) or more on a global severity of illness scale of 0–6. Patients in whom depressed mood was secondary to other predominant symptoms were excluded. Among such exclusions were other neurotic disorders, clear-cut schizophrenia, organic brain damage, alcoholism, or drug addiction accompanied by mild depression. The co-existence of personality disorder did not in itself comprise a ground for exclusion.

B. *Assessment measures*

The information for each study patient was collected from two research interviews. The initial interview took place at or shortly after admission when the patient was seen by the author and a semi-structured interview was conducted for history and clinical symptoms. Later, after clinical improvement or recovery, the patient was re-interviewed by a research assistant who gave a semi-structured interview covering life stress at the time of symptomatic onset of depression, and self-report personality scales were administered.

A number of assessment measures were included in the two interviews conducted.

(1) *Clinical interview for depression.* This scale was completed at the initial interview. It was based on an expanded version of the Hamilton Rating Scale for Depression (10). The items covered the symptomatic and historical variables important with respect to depressed patients. Most of the items were rated on 7-point scales, with closely defined anchor points. Although only one rater was used in this study an independent reliability study was conducted by the author and a second psychiatrist. The two raters made independent ratings on a different sample of patients interviewed by one or the other in the presence of both. This study indicated satisfactory reliability. The mean Pearson r between the two raters was 0·81 and there was agreement to within one point on 95 per cent of ratings.

(2) *Interviews for recent life events.* This semi-structured interview was administered to the patient by the research assistant at the second interview. It covered comprehensively the patient's life events occurring over the six months preceding the symptomatic onset of the depressive episode. It contained 60 events. Two scores were derived from this: (a) Total Recent Life Events Score (Total R.L.E. Score). This drew on work of Holmes and Rahe (14) who, by means of a scaling procedure, obtained weighted scores for a number of life events in terms of the social readjustment they necessitated. These scores were subsequently used to weight events occurring to a sample of subjects and the weighted scores were found useful in predicting the occurrence of medical illness (28). In the present study these weights were used, and the few additional items added to the list were arbitrarily assigned weights in keeping with the original list. Each event occurring to the patients was thus weighted by its score. The sum of these scores for the six month period was used as an index of total precipitating stress for each patient. (b) Based on narrative material recorded by the interviewer, the author also made for each patient a judgement of the

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

occurrence and degree of preceding stress. This was subsequently used as a validator. The total R.L.E. Score was found to be highly significantly related to it in the predicted direction.

(3) *The Maudsley Personality Inventory* (*M.P.I.*). Only the Neuroticism score from this self-report inventory was used in the present context, as an index of premorbid adjustment. The scale has been widely used, and evidence has been provided for its validity as an instrument to identify neurotic subjects (4). To minimize effects due to recent depression (2) the inventory was administered to patients at the second interview (after recovery) and instructions were given to the patient to answer according to usual state, not to that of recent illness (17). The whole inventory was administered although only the neuroticism score was used in these analyses.

C. *Method of cluster analysis*

The cluster analysis method used has been described by Friedman and Rubin (6). It is well suited to quantitative data and has been used in a study of borderline psychiatric patients (9). It has two special features. The first consists of the criterion for grouping. Any method for grouping must utilize some criterion measure to decide when a cluster is formed. The criterion adopted by this method uses covariance matrices (sums of squares and sums of cross products). The 'best partition' of the subjects is that which gives the highest value of the ratio of the determinant of the total covariance matrix to the determinant of the pooled within-groups covariance matrix (expressed as its logarithm, $\log_{10} \max \left|\frac{T}{W}\right|$). This is equivalent to partitioning the subjects in such a way as to minimize within-group generalized variance. The criterion measure cannot choose between different numbers of groups since it increases with the number of groups. The number of groups must be specified and settled ultimately on other criteria, although the step up in criterion value as groups are added may be a helpful indicator. A large increase suggests a worthwhile extra grouping. The criterion measure is identical with that used in linear discriminant function analysis where linear combinations of the original variables are derived to separate previously given groups in a space of reduced dimensionality. The difference is that in discriminant analysis the groups are given; in the cluster analysis we are to find the groups in such a way that they are maximally separated in the discriminant function space. As a result, discriminant function analysis also provides the best and most convenient axes on which to display the group separation.

The second special feature involves the computational methods. The procedure sets out to evaluate the effects of grouping individuals in all possible combinations to find that grouping which best satisfies the criterion. The total number of possible combinations of a moderate sized sample is immense, and to render the procedure viable a hill-climbing algorithm is applied (34). This involves computational shortcuts which render the procedure practical but non-exhaustive. There is some risk that the best partition may not be reached in a single analysis, and repeated analyses are therefore desirable.

D. *Computational procedures*

Prior to the cluster analyses a principal component analysis of the rating measures was carried out. Thirty-five variables were entered into this analysis. Twenty-nine of these were individual clinical symptom ratings, together with a global rating of severity of illness. The remainder were historical and premorbid variables including age, the number of previous depressions, a rating of previous excessive alcohol intake, length of illness, the M.P.I. neuroticism score, and the weighted stress score (Total R.L.E. score). From this analysis six principal components were obtained, the criterion for stopping being that the last component extracted should account for less than 10 per cent of the common factor variance. Loadings on these components have been described separately (26). The first appeared to be a general factor of severity and the second a bipolar factor contrasting psychotic (endogenous) and neurotic depression.

To ensure the statistical assumptions of normality and reduce data measures for computational economy, scores for the patients on

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

these six components, rather than the original 35 variables, were entered into the cluster analysis. They were scored as standard scores with mean of zero and standard deviation of one. A programme was used suitable for I.B.M. 360 computer (35). The analysis was carried out separately to isolate respectively two, three, four and five groups. To ensure that the best partition was reached the two-group analysis was completely run twice from independent random orders of patients; at the three-group level there were three separate analyses; at the four-group level, five; and at the five-group level, two. At each level the final best partition used for further evaluation was that of all replications giving the highest value of the criterion, log $\left|\frac{T}{W}\right|$. Reasons for preferring four groups will be presented.

A discriminant function analysis was carried out on the best partition at each level to display group separation. The original thirty-five variables were employed, rather than the six principal components. To aid in interpretation group mean scores on the thirty-five rating variables were also computed for the best partition at each level. In addition, relationships between groups at each level from two to five were examined as will be shown.

Results

A. *Criterion values*

Table I sets out the values of the partition criterion, log $\left|\frac{T}{W}\right|$ for the best partition at each level of grouping. There was a gradual increase in magnitudes as the number of groups increased. There was a substantial step of 0·82 from the two to the three level. The increase from three to four groups was a little less (0·67) and that from four to five groups less again (0·58). These steps were gradually diminishing and it did not appear that advancing beyond five groups would be profitable. There was no abrupt transition which might have indicated a preferred level of grouping.

Table I

Values of log $\left|\frac{T}{W}\right|$ for best partitions

Number of groups	2	3	4	5
Value of $\log_{10}\left\|\frac{T}{W}\right\|$	1·32	2·14	2·81	3·39

B. *Hierarchy of group structure*

Comparisons of group membership for the best partition at each level indicating a striking hierarchy of group structure. The comparisons were made by making cross tabulations of group membership from successive levels, e.g. two to three, three to four. These revealed that most individuals moved en bloc in clearly defined groups.

The hierarchical structure is summarized in Figure 1. Group labelling and interpretations will be presented further in due course. At the two-group level there were 66 members in one group and 99 in the other. At the three-group level the first group split into two new groups with 24 and 49 members, a few of them new. The second group remained substantially unchanged, with 92 members. Only nine individuals (5·6%) of the total 165 changed groups in a way not consistent with the hierarchical structure. In the change from three to four groups the remaining large group of 92 members split to two groups of 46 and 57. The other two groups showed little change. Only thirteen patients (7·9 %) changed groups in an inconsistent way. The two original groups had now each split to two new groups.

In the change from four to five groups the hierarchical structure was lost. The groups shuffled considerably and were joined by stray individuals so that with some of the new groups it was not entirely clear from where their core membership had come in the four-group system. At least 60 individuals (36·4 %) changed position across groups.

This hierarchy of groups was unexpected. Grouping at each level was a new procedure, and it was quite possible that very different organizations of the individuals at each level of grouping would give the highest value of the criterion measure. The hierarchy of structure provided strong evidence that a real structure and grouping of individuals was being demonstrated, rather than largely accidental and unstable groups.

The hierarchy of structure also suggested that four groups were the maximum that could usefully be derived. At the level of five groups the natural structure appeared to have become distorted by the constraint of having to form

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

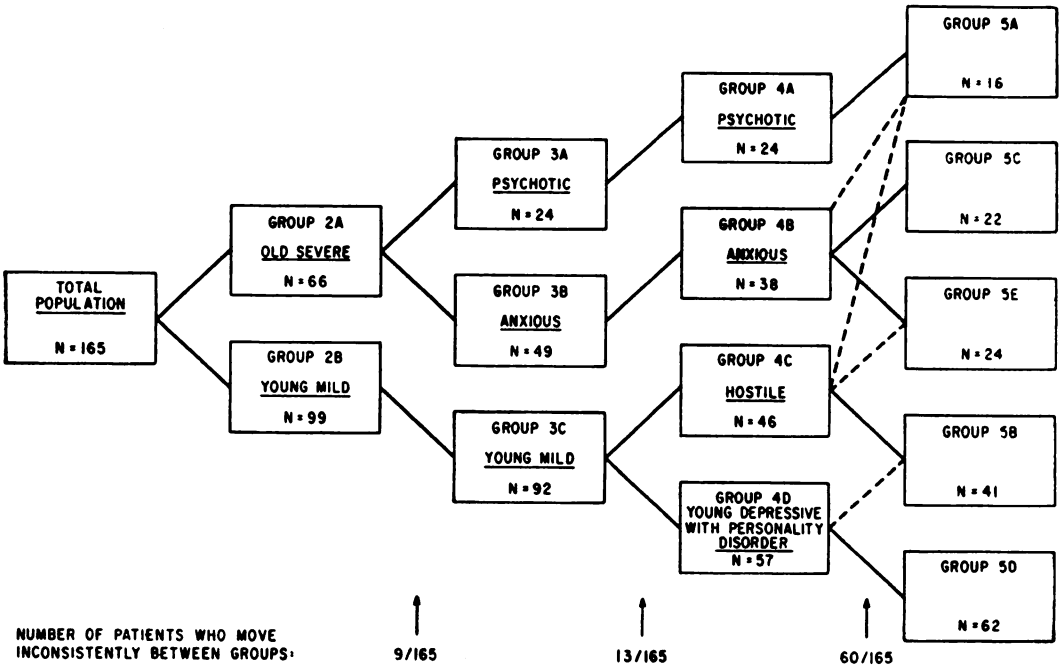


FIG. 1.—Hierarchy of groups.

five groups. Further support for this view came from examination of group means on the rating variables. Up to the four-group level it proved easy to make interpretations. At the five-group level interpretation was not possible. The five-group partition will therefore not be presented further.

C. *Group characteristics—two-group partition*

Set out in Table II are means for the 35 rating variables on the groups from the two-group partition. Differences were examined by one-way analysis of variance. The table shows F values but not significance levels. The F ratios validly reflect the extent to which different variables distinguish or fail to distinguish the groups. They do not, however, indicate true probability since the partitions, rather than being derived independently, were derived in such a manner as to produce maximal group separation on a combination of the rating variables.

Patients in Group 2A were older. In keeping with this they had experienced more previous episodes of illness. They showed less evidence of stress before illness, as judged by lower R.L.E. scores. M.P.I. neuroticism scores did not distinguish the two groups. Many symptoms did distinguish them and on almost all of these Group 2A scored higher. These differences were particularly apparent on ratings of core symptoms of depression which might reflect severity of illness, and the global rating of severity of illness.

In general, the axis separating these two groups appeared to be one of increasing age and severity of illness. This distinction shared some elements with the psychotic-neurotic or endogenous-reactive continua described in the literature, but the relationship was not close; for instance, the M.P.I. neuroticism scale, which might have been expected to distinguish the groups, did not do so. In addition, factor scores on the principal components were calculated and it was found that the two groups did not differ on the bipolar psychotic-neurotic factor.

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

TABLE II

Mean scores on rating measures for two-group partition

Items	Range	Group 2A Old severe N = 66	Group 2B Young mild N = 99	F value
		Mean	Mean	
Age	21–65	42·33	30·92	51·66
Length of illness (months)	*0–99	8·76	7·41	0·68
No. of previous depressions	*0–7	1·83	0·73	21·77
Excessive alcohol intake	0–3	0·05	0·17	0·27
Total RLE score	*0–402	77·53	134·96	20·78
MPI—neuroticism	0–48	34·73	33·96	0·15
Depressed feelings	0–6	4·26	3·58	24·33
Distinct quality	0–6	1·95	1·30	8·07
Diurnal variation (worse AM high)	0–12	7·34	5·27	28·55
Reactivity	0–6	2·61	3·24	13·07
Guilt, lowered self-esteem	0–6	3·06	2·34	17·11
Pessimism and hopelessness	0–6	3·17	2·54	11·92
Suicidal tendencies	0–6	3·41	3·31	0·11
Depersonalization	0–6	0·96	0·56	4·69
Obsessional symptoms	0–6	0·99	0·78	1·64
Helplessness	0–6	3·61	2·17	48·50
Work and interests	0–6	3·58	2·70	20·52
Energy and fatigue	0–6	3·55	2·80	8·19
Anxiety—psychic	0–6	3·52	2·23	54·16
Anxiety—somatic	0–6	3·11	2·02	29·70
Constipation	0–6	1·20	0·36	15·69
Change of appetite (anorexia high)	0–12	9·38	7·28	27·59
Hypochondriasis	0–6	1·86	0·79	21·45
Paranoid ideas	0–6	0·54	0·79	1·53
Irritability	0–6	1·38	2·71	31·70
Initial insomnia	0–6	1·88	2·30	1·91
Middle insomnia	0–6	2·35	2·04	0·98
Delayed insomnia	0–6	2·58	1·01	32·08
Loss of insight	0–6	1·17	0·63	6·04
Retardation	0–6	0·69	0·42	2·91
Agitation	0–6	1·77	0·71	31·10
Hostility	0–6	0·73	0·70	0·03
Self-pity	0–6	1·22	0·90	2·17
Depressive delusions (no. of patients)	—	8	2	—
Global rating of severity of illness	0–6	3·65	2·99	23·55

* Actual range of scores. For remaining items, range given is range of scale used.

D. *Group characteristics—four-group partition*

The full group structure was displayed in the four-group partition, of which mean scores on the rating variables are shown in Table III. For the same reasons as previously significance levels are not given. Since in the usual circumstances the 5% F value for 3, 16 d.f. would be 2·65, these groups were highly differentiated on almost all variables. The groups labelled 4A and 4B in the table were derived from Group 2A and separated at the three-group level. Groups 4C and 4D were derived from the original Group 2B and did not split until the four-group level.

The older patients were divided into two contrasting and readily interpretable groups. Group 3A contained nine of the ten patients included in the analysis who showed depressive

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

TABLE III

Mean scores on rating measures for four-group partition

Items	Range	*Group 4A* Psychotic N = 24	*Group 4B* Anxious N = 38	*Group 4C* Hostile N = 46	*Group 4D* Young Depressive with Personality disorder N = 57	F value
		Mean	Mean	Mean	Mean	
Age	21–65	46·83	39·74	33·53	29·25	21·22
Length of illness	*0–99	6·58	11·90	5·73	7·57	2·81
No. of previous depressions	*0–7	1·00	2·42	0·91	0·63	12·82
Excessive alcohol intake	0–3	0·46	0·31	0·52	0·33	0·68
Total RLE score	*0–402	96·81	76·89	140·49	121·49	4·98
MPI—neuroticism	0–48	25·93	38·65	29·81	37·02	8·38
Depressed feelings	0–6	3·67	4·58	3·76	3·54	12·65
Distinct quality	0–6	2·47	1·97	1·43	1·10	6·17
Diurnal variation (worse AM high)	0–12	6·79	7·84	5·58	5·04	11·83
Reactivity	0–6	2·42	2·58	2·96	3·52	8·77
Guilt, lowered self-esteem	0–6	3·17	2·97	2·16	2·57	6·06
Pessimism and hopelessness	0–6	3·38	2·95	2·87	2·41	4·40
Suicidal tendencies	0–6	2·96	3·63	3·62	3·13	1·25
Depersonalization	0–6	0·46	1·34	0·40	0·68	5·48
Obsessional symptoms	0–6	0·79	1·24	0·51	0·93	3·82
Helplessness	0–6	3·29	3·53	2·11	2·57	9·00
Work and interests	0–6	3·67	3·61	3·00	2·54	8·51
Energy and fatigue	0–6	2·75	4·08	3·09	2·58	7·05
Anxiety—psychic	0–6	3·29	3·55	1·87	2·71	18·90
Anxiety—somatic	0–6	2·29	3·58	1·62	2·46	19·17
Constipation	0–6	1·58	1·08	0·60	0·13	8·41
Change of appetite (anorexia high)	0–12	9·48	9·19	7·27	7·45	7·27
Hypochondriasis	0–6	2·38	1·87	0·93	0·55	12·89
Paranoid ideas	0–6	1·42	0·46	0·67	0·57	3·32
Irritability	0–6	1·67	1·24	2·44	2·86	10·10
Initial insomnia	0–6	2·25	2·00	2·60	1·88	1·29
Middle insomnia	0–6	2·21	2·05	2·41	1·98	0·43
Delayed insomnia	0–6	2·96	2·05	0·84	1·36	8·62
Insight	0–6	3·17	0·46	0·82	0·14	54·18
Retardation	0–6	1·09	0·50	0·79	0·11	7·40
Agitation	0–6	2·29	1·39	0·70	0·81	11·82
Hostility	0–6	1·58	0·42	1·20	0·11	16·83
Self-pity	0–6	0·79	1·46	1·40	0·57	5·10
Depressive delusions (no. of patients)	—	9	1	0	0	—
Global rating of severity of illness	0–6	4·42	3·45	3·11	2·77	29·02

* Actual range of scores. For remaining items, range given is range of scale used.

ideas such as guilt or hopelessness, of delusional intensity. Patients in this group had the highest scores for global severity of illness and highest scores on many individual variables. These included such typically psychotic or endogenous symptoms as distinct quality, guilt, anorexia, retardation and delayed insomnia. They were the oldest patients. They had low scores on the total R.L.E. score and strikingly low scores on the M.P.I. neuroticism scale. They may reasonably be regarded as psychotic depressives.

The second group of patients, Group 4B, showed different features. They showed the highest scores on psychic and somatic anxiety, depersonalization, obsessional symptoms and fatigue. In addition to this admixture of anxiety

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

and neurotic symptoms, they were moderately depressed. They had lower scores on the global illness scale than the psychotic group, but the highest scores on depressed feelings and suicidal tendencies. They had the greatest number of previous episodes of depression. They had the highest scores on M.P.I. neuroticism, but low scores on the total R.L.E. score. The patients may be tentatively labelled anxious depressives. They appear to be a neurotic group but cannot be regarded as reactive since they have relatively low stress scores.

The two groups of younger patients were Groups 4C and 4D. The separation between them was less striking and they shared many features. Group 4C were a little more ill than Group 4D on most symptom scores. They had particularly high scores on hostility and self-pity. The psychotic group had similar scores on these variables, but in this much less severely ill group the hostility and self-pity were striking in comparison with the general mild symptomatology. They could therefore be regarded as hostile depressives.

Group 4D consisted of the youngest patients. They had the highest scores on M.P.I. neuroticism, and relatively high scores on the total R.L.E. score. On clinical symptoms they were at the mild extreme, with low scores on most individual symptoms and the global severity rating. They showed particularly high scores on the rating for reactivity, indicating fluctuations in the depressed mood related to immediate environmental changes. Subsequent examination of case records indicated a particularly high incidence of disturbed life pattern among these patients and a tendency to receive additional routine clinical diagnoses of personality disorders. They were therefore labelled as young depressives with personality disorder.

Case records and narrative summaries prepared for the study on each patient were inspected. They supported the interpretations made for the four groups. They also revealed some overlap between groups, particularly between Group 4C and Group 4D.

The results had suggested the two older groups were comparable in amount of preceding stress, contrary to what might have been predicted from the endogenous-reactive dichotomy. This finding depended on the weighted total R.L.E. score. It was therefore checked against judgements made by the author on the presence or absence of a precipitant, based on narrative summaries of the interview for recent life events. In terms of this judgement 20 per cent of Group 4A and 36·1 per cent of Group 4B appeared to have unprecipitated episodes. For Group 4C the corresponding figure was 7·7 per cent and for Group 4D, 4·4 per cent. On this criterion, therefore, both psychotic and anxious groups also reported comparable stress; the incidence appeared a little less in the anxious patients. Moreover, any distinction between the four groups based on stress was weak since, even in psychotic and anxious groups, the majority of patients did experience at least some possible precipitant events prior to illness.

E. *Plot of individuals on discriminant axes*

Discriminant function analyses were carried out on the partitions at each level, using the original 35 variables rather than the six principal components. Individuals were then plotted by their scores on the resultant canonical variates (29). Fig. 2 displays the position of individuals from the four-group partition. There were three canonical variates; their associated eigenvalues were 4·37, 2·64 and 0·70. Most

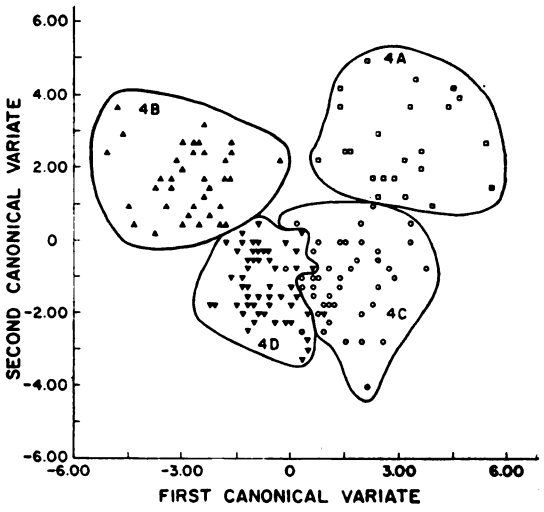


FIG. 2.—Four group partition. Plot of patients on first two canonical variates.

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

separation was therefore on the first two axes, which were plotted. They indicated four approximate clusters. Separation was poorest between Groups 4C and 4D, although these two groups were also separated to some degree on the third variate, not shown in the diagram.

It should be noted that no firm conclusions about group separation can be based on this discriminant function analysis. The usual tests of significance for the discriminant functions do not apply and the probability theory associated with the method has not been worked out (6). The sole criterion for forming groups was maximal separation on the 35 variables as represented by their factor condensations. The discriminant functions which display that separation maximally are based on the same variables and cannot directly test the separation.

Discussion

This study has suggested that depressed patients may be usefully classified into four groups. These groups form a hierarchical structure in which the first level divides patients into two groups: one group of older patients more severely ill and approximating to classical melancholics, and one group younger and more mildly ill. At the next level each of these higher order groups splits into two. The evidence that four groups comprise the maximum number that can usefully be obtained rests primarily on the hierarchical structure and on conceptual meaning, both of which are lost in the transition from four to five groups, apparently because the constraint of forming five groups overcomes the natural structure in the data.

The four groups revealed by the full schema cover a wide range of heterogeneity. They rank approximately in order of mean age and severity of illness. At one extreme were psychotic depressives, oldest and distinguished by marked severity of illness, often to the point where delusions of depressive type were present. These patients showed a variety of features, such as a distinct and different quality to the depression, guilt, anorexia, retardation, and delayed insomnia, which have been associated with psychotic and endogenous depression in the literature. They had relatively low scores on M.P.I. neuroticism, indicating a non-neurotic premorbid adjustment, and showed less evidence of stress at the onset of illness although in the majority of cases some stress was present.

The second group of patients was moderately severely depressed. Their symptoms were characterized by a strong admixture of anxiety and a greater tendency than other groups to show additional neurotic manifestations such as obsessional symptoms and depersonalization. These patients tended to be a little younger than the psychotic depressives; they had, nevertheless, experienced the greatest number of previous episodes of illness. They were particularly characterized by high scores on M.P.I. neuroticism. They were not distinguished from the psychotic depressives by preceding stress. They have been labelled as anxious depressives; they might alternatively be regarded as middle aged neurotic depressives.

The third and fourth groups of patients showed some overlap. One group tended to be older and rather more depressed. Their most characteristic feature was a strong admixture of hostility and self-pity. The last group of patients was the youngest. Symptomatically their illnesses were mild and symptoms fluctuated moderately in response to environmental events. These patients had high neuroticism scores and life histories showing disturbed interpersonal relations suggestive of personality disorder. Almost all had experienced stress prior to the onset of illness. Although the depressive syndrome was milder in both these groups, it was pathological to the point where it fitted the study criteria in intensity and was the predominant symptomatic abnormality. Mean scores for depressed feelings were in the moderate range and for suicidal tendencies as high as in the other groups.

Comparison with other studies

Some other studies have suggested groups which overlap with the present ones. Three of the four patterns described by Grinker *et al.* (8) appear similar. Their Pattern A, characterized by feelings of dejection, hopelessness, and low self-esteem, together with retardation, shares some features with the psychotic group. Patients of Pattern B, who showed considerable feelings of depression together with marked anxiety,

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

appear very similar to the anxious depressives. Pattern D was characterized by provocative, demanding and hostile behaviour, similar to that shown by the present hostile depressives. Pattern C, characterized by agitation, demanding behaviour, and hypochondriasis, does not however correspond to the young depressives with personality disorder.

There appear to be particularly close parallels to the three groups described by Overall *et al.* (24). These authors carried out repeated factor analyses of patient profiles using patient samples derived from Veterans' Administration hospitals. Ratings were based on the Brief Psychiatric Rating Scale which covers a broader spectrum of symptoms than were rated in the present study but with fewer specifically depressive symptoms and no historical variables. Three groups of patients were described: retarded, anxious, and hostile. The retarded group appear to resemble the psychotic patients in the present study, who scored particularly high on retardation. Parallels between the two sets of anxious and hostile groups are evident.

Two other studies employing factor analysis proposed groups resembling at least one of the four groups. Hamilton and White (11) interpreted one factor as reflecting 'psychopathic' depression, or depression in an abnormal personality. Factor loadings indicated such patients should be younger and characterized by an absence of the following: guilt, delayed insomnia, loss of interest, weight loss, and gastrointestinal symptoms. This description suggests similarities to the young depressives with personality disorder. Rosenthal and Gudeman (32) reported a self-pitying constellation characterized by self-pity, blaming the environment rather than self, hypochondriasis, anxiety, demanding and complaining behaviour, reactivity, irritability, hostility, and absence of both guilt and retardation. This overlaps with the characteristics of the hostile group.

The young depressives with personality disorder have received little attention except from the Hamilton and White study. Most research study samples have been confined to inpatients. The present sample included a large outpatient group among whom these patients were particularly represented (25). This may well explain their clear emergence here. Since under modern conditions the majority of depressed patients are treated as outpatients this group of young patients whose depression develops in a background of personality disorder are likely to prove increasingly important. They probably represent a historical trend; several authors have documented a gradual historical change in the characteristics of depressed patients presenting for treatment, in the direction of younger, more mildly ill and more neurotic patients predominating (22, 23, 30).

One postulated diagnostic grouping which was not reflected in any partition group was involutional melancholia. These findings are consistent with other recent studies which have failed to demonstrate any grounds for regarding it as a separate diagnostic entity (3, 16).

Comparison with two-type classification of depression

The four-group classification is consistent with the psychotic-neurotic dichotomy which predominates in the literature, but goes beyond it. The first division into two groups did not correspond exactly to this dichotomy, although its main elements appeared to be age and severity of illness. The four-group partition made the situation clearer. One of the two older groups corresponded clearly to psychotic depressives. The remaining three groups could all be regarded as neurotic, in the sense of absence of delusions, in each displaying some symptoms said to characterize neurotic depressive groups (20) and by evidence of poor premorbid adjustment as evidenced by disturbed life histories and by high scores on the M.P.I. neuroticism scale. Comparison with M.P.I. norms must be made cautiously, since age, sex and cultural differences from the published English samples, and residual effects of depression, may all interfere with scores. The norm for the neuroticism score is 19·89 (4). Only the mean score for the psychotics (25·93) approached at all near it. The remaining groups all had scores in the neurotic range of standardization scores—38·65 for anxious depressives, 37·02 for young depressives with personality disorder and 29·81 for hostile depressives.

It thus appears that the two-type theory, although useful, fails adequately to classify the diversity apparent among the non-psychotic

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

depressives. These have always received less attention in the literature, attention having been focused on the psychotic or endogenous patients for whom a possible biological aetiology makes physiological and biochemical investigations attractive. General experience suggests that the heterogeneity of depressed patients is considerable. If it were regarded as sufficiently systematic to enable classification in the first place it was to be expected that two polar types would fail to do full justice to this diversity.

Two dichotomies, endogenous-reactive and psychotic-neurotic, have been loosely used as equivalent labels in describing the two types. Although no labels are ideal, and semantic debates do not greatly illuminate this field, the results suggest that the psychotic-neurotic terminology may be preferable. It well describes the severity of illness characterizing one group in which many of the patients (although somewhat less than half) even fit a narrow definition in the sense of manifesting delusions. It also describes the disturbed life history or neuroticism apparently characterizing each of the remaining three groups. Life stress implied in the endogenous-reactive concept was less effective in distinguishing these predominantly symptomatic groups. It was common in all patients. Although less common for the psychotic patients it was even a little less frequent among one neurotic group—the anxious patients. Garmany (7) also found precipitants common among depressives diagnosed as endogenous.

Implications for methodology

The results of this study exemplify some important issues for the methodology of classification. Previous factor analytic studies have suggested a simple division of depressives into two polar types, rather than more complex classifications. The principal component analysis carried out on the present data revealed one factor interpretable as indicating severity, another corresponding to the psychotic-neurotic continuum, and a third apparently contrasting anxiety with typical depressive symptoms (26). Three more factors were not interpretable. All factors except the first were weak. Plots of individuals did not show any clustering. There was little hint of the coherent structure of four interpretable groups found on cluster analysis.

These present results illustrate the greater suitability of cluster analysis grouping techniques for the purpose at hand. Fundamentally, factor analysis of rating measures is not intended to classify individuals and it has serious limitations in this respect. One limitation has already been referred to. Factors are dimensional in nature and usually multiple and orthogonal, so that variation along each is independent. Description of position on such a dimension says nothing as to position on any of the others, and has very different implications from membership in one of a set of mutually exclusive groups. The awkwardness of such interpretations particularly obtrudes if multiple factors are used to argue for multiple groups. Moreover, although some factors may emerge which parallel groups, they need not do so (21). This may occur, for instance, if there are several groups in which the patterns of intercorrelations are different, so that they are likely to be obscured in the total population.

Factors do form legitimate dimensions on which to plot the position of individuals to examine for the bimodality of distribution or visual clustering that would indicate a distinct group (21). Even here there are serious problems. We are limited to working in three dimensions at any time, and in most cases to the two dimensions of a sheet of paper. Much information must be ignored. If the grouping is weak, or none of the dimensions of the factor space lies across the best axis for separation, potential clusters will be missed.

In contrast, the cluster analysis technique employed here is specifically intended to isolate groups of individuals and is capable of working in multidimensional hyperspace to do so. It is a very powerful technique. Criticism of factor analysis should not be overstated; it has undeniably proved useful in the classification of depression and its introduction helped give a new impetus to fruitful research in this much debated field. Techniques for grouping individuals by cluster analysis, also referred to as numerical taxonomy (36), are newer and less well worked out. The method used in this study is only one of a number available. It cannot choose mathematically between different num-

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

bers of groups. Like all statistical techniques for classifying it is more an exploratory tool to suggest possible classifications than one suitable for formal hypothesis testing. Nevertheless it is desirable that future studies concerned with psychiatric classification and grouping by mathematical methods should make increased use of such clustering techniques.

Group boundaries and the nature of groups

The plot of individuals in Fig. 2 suggested relatively clear boundaries with minimal overlap between groups. This is misleading; the axes derived from discriminant function analysis are best designed to display the separation that has been obtained by grouping individuals using the sole criterion that separation in terms of the variables should be maximal. Like other multivariate techniques this one is very powerful, and makes group assignment in a complex way based on the weighting of many variables without giving any particular priority to those that may appear conceptually central. There are three grounds for regarding boundaries between the four groups as at best very approximate. First, the plots compare unfavourably with the tight clusters separated by wide spaces that can be obtained using this method on well-structured data (6). Secondly, when alternative partitions obtained from the repeat analyses were examined at each level of grouping they were often found to involve considerable changes of group membership. These alternative partitions were all mathematically inferior to the best partitions utilized, but in some cases only very little so. That relatively large changes would occur suggests that in spite of the clear structure group boundaries are approximate. Thirdly, both the scores on rating measures and subsequent inspection of case histories and narrative summaries revealed that, although the group characteristics were apparent, there was moderate variation within groups and moderate overlap between groups.

In these circumstances the groups must be regarded more as stereotypes than precise categories with clear cut-off points. With an appropriate method the weak and approximate boundaries that exist in the data can be detected. The difficulty in making clear boundaries between diagnostic groups in psychiatry is well known and is not surprising in view of the multifactorial aetiology and multiplicity of pathoplastic factors determining symptomatic picture in any case. This situation contrasts with the much more clear-cut aetiology and pathology pertaining to general medicine, from which the disease model derives. It cannot be argued that the present groups escape the basic problem in grouping depressed patients; many patients lie near group boundaries, and precisely which side of the boundary is uncertain.

In this situation, Kendell (16) has argued for the abandonment of such diagnostic subcategories, and their replacement by dimensions such as the psychotic-neurotic continuum, to which factors correspond. This approach may have great value, particularly for research purposes such as predictive studies. Dimensional analysis allows much more precise description than may be possible in forced assignment to a group. For general purposes, however, it has its limitations. The precision is not obtainable in ordinary clinical practice since it depends on reliable rating methods and procedures for combining and weighting items which are extremely time-consuming in the absence of computer facilities. More importantly, as the number of axes increases the schema soon becomes extremely cumbersome and the relationship of individuals to each other becomes impossible to appreciate.

For these reasons it seems likely that psychiatric classification will continue to be based not on dimensions but on the familiar model of diagnostic categories to which individual patients are assigned and to which the present four groups correspond. Within a major syndrome like depression such categories are likely to have only approximate boundaries, so that many individuals may be difficult to assign. For research purposes a procedure is available to assign them —the discriminant function. For general purposes the clinician's 'best guess' is likely to remain the usual procedure. Definitions must also be approximate. It is arguable whether specific aetiological or other features will ever be found to underlie syndromes such as these. At present it is not clear what such specific underlying

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

features are. Any which might be proposed, such as neuroticism or precipitant stress, themselves seem continuous dimensions without clear cut-off points. Therefore, we cannot use simple logical definitions of the syndromes based on key features (such as a simple endogenous-reactive dichotomy based solely on precipitant stress). Syndromes such as those dealt with here can only be defined loosely in terms of clusters of features, and individuals assigned approximately in terms of degree of resemblance of the various features that they show. The present groups at least reflect the weak natural boundaries suggested by the patients themselves in their scores on the rating measures.

The philosophy underlying attempts at psychiatric classification has been reviewed by Hempel (12). Essentially, diagnosis in this context involves classification of persons manifesting abnormal behaviour at one point in time. Such classification facilitates communication and may be of value in providing conceptual simplification, explanatory theories and predictive information regarding outcome or response to treatment. The more information that can be generalized from a classificatory schema the greater is its value.

The evidence for the groups suggested by this study has not gone beyond a very preliminary stage. The only justifications that can be offered for the groups are that they comprise the natural categories and boundaries indicated mathematically by the patient ratings themselves, that they are conceptually meaningful, and that they resemble other groupings described in the literature. They as yet require replication in other samples, and validation against other variables such as biological or psychological correlates, follow-up outcome, or response to treatment. Such studies are under way. For the meantime the groups must be regarded as exploratory. As methods for treatment of depression become more plentiful and diverse, differential assignment of the best treatment is likely to become increasingly important. In these circumstances the ability approximately to systematize the heterogeneity of depressed patients, which is implied by any type of classification, may acquire increasing practical utility.

Summary

One hundred and sixty-five depressed patients from a variety of treatment settings were rated as to clinical symptoms, previous history, life stress, and premorbid neuroticism, and the ratings were subjected to a special cluster analysis for classifying individuals.

The results indicated a hierarchical structure of groups. At the first level there were two groups, one older and more severely ill, one younger and more mildly ill. Subsequent division of each of these produced four approximate groups. One group was characterized by a severe illness, sometimes with delusions, on a background of good premorbid adjustment, and corresponded to psychotic depressives as described in the literature. A second group consisted of moderately depressed patients with a strong admixture of anxiety, a high incidence of previous illnesses and high neuroticism scores. A third group comprised depressives with a considerable element of hostility; and the last group contained young patients whose relatively mild illnesses developed on a background of personality disorder.

These groups resemble other groups of depressed patients previously described in the literature. They are also consistent with the psychotic-neurotic dichotomy, but suggest that non-psychotic depressives are diverse and comprise several groups. The emergence of these multiple and easily interpretable groups of depressed patients also suggests the utility of such cluster analysis grouping techniques for studying the classification of psychiatric patients.

Acknowledgements

This study was supported in part by U.S.P.H.S. Grant MH 13738 from the National Institute of Mental Health, Unites States Department of Health, Education and Welfare. I am grateful to Dr. Gerald L. Klerman and Mrs. Brigitte Prusoff for advice and collaboration at all stages of the study, and to Mr. Herman Friedman for his advice. The analysis was carried out by Mrs. Prusoff on the IBM 360 computer of the Information Research Division of Rockland State Hospital, New York. The material of this paper forms part of a thesis which has been submitted to the University of Otago, New Zealand.

References

1. Carney, M. W. P., Roth, M., and Garside, R. F. (1965). 'The diagnosis of depressive syndromes and the prediction of ECT response.' *Brit. J. Psychiat.*, **111**, 659–74.

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press

2. COPPEN, A., and METCALFE, M. (1965). 'Effect of a depressive illness on MPI scores.' *Brit. J. Psychiat.*, **111**, 236–9.
3. DALY, R. J., and COCHRANE, C. M. (1968). 'Affective disorder taxonomies in middle aged females.' *Brit. J. Psychiat.*, **114**, 1295–7.
4. EYSENCK, H. J. (1959). *Manual of the Maudsley Personality Inventory*. London.
5. FAHY, T. J., BRANDON, S., and GARSIDE, R. F. (1968). 'Clinical syndromes in a sample of depressed patients: a general practice material.' *Proc. Roy. Soc. Med.*, **62**, 331–5.
6. FRIEDMAN, H. P., and RUBIN, J. (1967). 'On some invariant criteria for grouping data.' *J. Amer. Stat. Assn.*, **62**, 1159–78.
7. GARMANY, G. (1958). 'Depressive states: their aetiology and treatment.' *Brit. med. J.*, **2**, 341–4.
8. GRINKER, R. R., MILLER, J., SABSHIN, M., NUNN, R., and NUNNALLY, J. C. (1961). *The Phenomena of Depressions*. New York.
9. GRINKER, R. R., WERBLE, B., and DRYE, R. C. (1968). *The Borderline Syndrome: A Behavioural Study of Ego Functions*. New York.
10. HAMILTON, M. (1960). 'A rating scale for depression.' *J. Neurol. Neurosurg. Psychiat.*, **23**, 56–62.
11. HAMILTON, M., and WHITE, J. M. (1959). 'Clinical syndromes in depressive states.' *J. ment. Sci.*, **105**, 988–98.
12. HEMPEL, G. (1961). 'Introduction to problems of taxonomy.' In *Field Studies in Mental Disorders*. (Ed. J. Zubin.) New York.
13. HOLLISTER, L. E., OVERALL, J. E., SHELTON, J., PENNINGTON, V., KIMBELL, I., and JOHNSON, M. (1967). 'Drug therapy of depression: amitriptyline, perphenazine and their combination in different syndromes.' *Arch. gen. Psychiat.*, **17**, 486–93.
14. HOLMES, T. H., and RAHE, R. H. (1967). 'The social readjustment rating scale.' *J. psychosom Res.*, **11**, 213–8.
15. KAY, D. W. K., GARSIDE, R. F., BEAMISH, P., and ROY, J. R. (1969). 'Endogenous and neurotic syndromes of depression: a factor analytic study of 104 cases. Clinical features.' *Brit. J. Psychiat.*, **115**, 377–88.
16. KENDELL, R. E. (1968). *The Classification of Depressive Illnesses*. Maudsley Monograph No. 18. London. Oxford University Press.
17. KENDELL, R. E., and DISCIPIO, W. J. (1968). 'Eysenck Personality Inventory scores of patients with depressive illnesses.' *Brit. J. Psychiat.*, **114**, 767–70.
18. KILOH, L. G., and GARSIDE, R. F. (1963). 'The independence of neurotic depression and endogenous depression.' *Brit. J. Psychiat.*, **109**, 451–63.
19. LYERLY, S. B. (1965). 'A survey of some empirical clustering procedures.' In: *The Role and Methodology of Classification in Psychiatry and Psychopathology*. (Eds. Katz, M. M., *et al.*) Public Health Service Publication No. 1584. U.S. Dept. of Health, Education and Welfare, Bethesda, Maryland.
20. MENDELS, J., and COCHRANE, C. (1968). 'The nosology of depression: the endogenous-reactive concept.' *Amer. J. Psychiat.*, **124**, 1–11.
21. MORAN, P. A. P. (1966). 'The establishment of a psychiatric syndrome.' *Brit. J. Psychiat.*, **112**, 1165–71.
22. MUNCIE, W. (1963). 'Depression or depressions?' *Canad. psych. Assn. J.*, **8**, 217–24.
23. OLTMAN, J. E., and FRIEDMAN, S. (1965). 'Trends in admissions to a state hospital, 1942–1964.' *Arch. gen. Psychiat.*, **13**, 544–51.
24. OVERALL, J. E., HOLLISTER, C. E., JOHNSON, M., and PENNINGTON, J. (1966). 'Nosology of depression and differential response to drugs.' *J.A.M.A.*, **195**, 946–8.
25. PAYKEL, E. S., KLERMAN, G. L., and PRUSOFF, B. (1970). 'Treatment setting and clinical depression.' *Arch. gen. Psychiat.*, **22**, 11–21.
26. PAYKEL, E. S., PRUSOFF, B., and KLERMAN, G. L. 'The endogenous-neurotic continuum in depression: rater independence and factor distributions.' *J. psychiat. Res.* In press.
27. PILOWSKY, I., LEVINE, S., and BOULTON, D. M. (1969). 'The classification of depression by numerical taxonomy.' *Brit. J. Psychiat.*, **115**, 937–45.
28. RAHE, R. H. (1968). 'Life-change measurement as a predictor of illness.' *Proc. Roy. Soc. Med.*, **61**, 1124–8.
29. RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York.
30. ROSENTHAL, S. H. (1966). 'Changes in a population of hospitalized patients with affective disorders.' *Amer. J. Psychiat.*, **6**, 671–81.
31. ——, and GUDEMAN, J. E. (1967). 'The endogenous depressive pattern: an empirical investigation.' *Arch. gen. Psychiat.*, **16**, 241–9.
32. ——, and —— (1967). 'The self-pitying constellation in depression.' *Brit. J. Psychiat.*, **113**, 485–9.
33. ——, and KLERMAN, G. L., (1966). 'Content and consistency in the endogenous depressive pattern.' *Brit. J. Psychiat.*, **112**, 471–84.
34. RUBIN, J. (1967). 'Optimal classification into groups: an approach for solving the taxonomy problem.' *J. theoret. Biol.*, **15**, 103–44.
35. ——, and FRIEDMAN, H. P. (1967). *A Cluster Analysis and Taxonomy System for Grouping and Classifying Data*. IBM Computer Program. New York.
36. SOKAL, R. R., and SNEATH, P. H. A. (1963). *Principles of Numerical Taxonomy*. San Francisco.

E. S. Paykel, M.B., M.R.C.P., M.R.C.P.Ed., D.P.M., *Assistant Professor, Department of Psychiatry, Yale University School of Medicine, 34 Park Street, New Haven, Connecticut, 06519, U.S.A.*

(Received 9 December 1969)

https://doi.org/10.1192/bjp.118.544.275 Published online by Cambridge University Press