Brit. J. Psychiat. (1981), **139**, 357–362

Comments

HAVE MULTIVARIATE STATISTICS CONTRIBUTED TO CLASSIFICATION?

In the last twenty years many workers have applied multivariate statistical techniques to problems of classification in psychiatry. To what extent have the techniques resolved the problems? This personal view will attempt to appraise the current situation.

Psychiatric classification is an offshoot of medical classification. In medicine disorders are definable at two levels: (i) the syndrome, or constellation of symptoms with a common core pathophysiology, e.g. the malabsorption syndrome, bronchopneumonia, heart failure; (ii) the disease, with a single predominant cause (there may be other predisposing and contributing factors). There may be one syndrome associated with a disease, or several if different anatomical, biochemical, physiological loci are affected. These syndromes may be unique, or shared with other causes.

In general medicine each disease is regarded as having specific well defined aetiology, pathology, clinical picture and treatment. Diseases are thought of as clear-cut entities, either present or absent, uncommonly multiple and rarely overlapping. The time-honoured way of isolating disorders has been first to separate off syndromes on the basis of clinical features, then to investigate pathological mechanisms and causes underlying them. The advantages of the diagnostic system have been very great. Once a patient is correctly diagnosed, much more information can be applied to him from the general body of knowledge, regarding causation, outcome and treatment (Hempel, 1961).

In psychiatry the development of a satisfactory nosology based on these models has had only limited success (Kendell, 1975). Some have questioned whether the disease model is appropriate. It has been difficult to isolate clear-cut diseases with unified aetiology, pathology, clinical manifestations and treatment. There has been some success in delineating very broad categories such as schizophrenia and depression. Even here boundaries may be weak and mixed cases common. These disorders correspond more to syndromes, defined in symptom terms with multiple aetiologies, than to diseases. Finer diagnostic categories such as sub-types of depression have emerged in an even less clear-cut way, and it may be difficult to accept the word of the clinician who claims to descry them, as reflecting anything more than his untempered enthusiasm.

One way around this problem is to separate groups on the grounds of a single criterion which is clear-cut, and then to demonstrate other differences between them. The bipolar-unipolar dichotomy in depression is a good example of this.

Suppose, however, there is no single clear-cut criterion. Perhaps no single criterion exists; multifactorial aetiology is common in psychiatry, even in the single case. Perhaps there is an underlying biochemical, physiological or psychological distinction which we have not yet discovered, but which is only weakly related to the symptoms and other superficial features accessible to us.

Here the psychiatrist may turn, as desperate men will, to statistics, and particularly to multivariate techniques. Multivariate methods use several variables simultaneously. By their nature they do not assign a priori weight to any special variable. This is the very antithesis of defining a classification by a single key concept. It can be very annoying to those who like their knowledge to be clear-cut. On the other hand it provides tools which can be subtle and flexible, provided that any interpretations are made with caution.

The profusion of papers using these techniques, their complexity, impressive statistical plumage, and the vigorous scholastic debate between proponents of different schools, have understandably put off many psychiatrists. There is a risk that the two schools of psychiatry, statistical and clinical, may fall totally out of communication. In fact, although the statistical techniques are complex, the principles behind what they are trying to do are fairly simple.

The techniques

There are three main multivariate techniques which have been used in psychiatry: factor analysis, cluster analysis and discriminant analysis. Their use has also been discussed by Kendell (1975) and Garside and Roth (1978). Each has a different aim corresponding roughly to a different level in the classificatory debate. As in the clinical method, the aim is first to group symptoms into syndromes, and then to group individuals with syndromes into diagnostic categories.

Factor analytic techniques, including principal

https://doi.org/10.1192/bjp.139.4.357 Published online by Cambridge University Press

component analysis, are essentially dimensional. They examine relationships between variables, usually from a matrix of inter-correlations, and simplify them to a smaller number of hypothetical dimensions, which can be supposed to summarize the main axes of actual variation in the data. Each dimension is a continuous measure, like height. In most techniques the dimensions are orthogonal (i.e. at right angles), so that variation along one is independent of variation along the others.

The dimensional nature of factor analysis makes it particularly suitable for simplifying rating instruments into smaller sets of multi-item measures, or dealing with dimensional concepts such as intelligence, through which it first appeared in psychology. It is also suitable for measuring the disturbances which go to make up syndromes: the emergence of a factor suggests that the features in it relate together and this can be difficult to absorb on simply inspecting a complex matrix of inter-correlations, without simplification.

A rudimentary form of factor analysis was the first multivariate technique applied to psychiatric classification when Moore (1930) sought syndromes among data on symptoms, history and psychological tests. Wittenborn *et al* (1953) derived nine dimensions in psychotic illness. The first factor analytic study on depressive classification was by Hamilton (Hamilton, 1960; Hamilton and White, 1959).

Because factor analysis produces dimensions it is not ideally suited to the usual question in psychiatric classification—whether individuals characterized on symptoms and other features at a time of disorder form separate groups corresponding to different diagnoses. Since factors are continuous dimensions and orthogonal, to say that an individual scores high on a specific factor says nothing as to his scores on other factors, a situation very different to membership of one of a set of mutually exclusive groups. This is the usual conception of diagnosis, and although it has been argued that psychiatric disorders may be dimensional (Kendell, 1968), it is not in practice easy for clinicians to think in such terms or to characterize the position of individuals simultaneously on four or more dimensions. Factor analysis is not really suitable for grouping individuals. A variant, employing q correlations (correlations between individuals on a set of measures) has sometimes been used for this purpose, but it too produces dimensions.

In recent years techniques of cluster analysis or numerical taxonomy have been derived specifically to form groups of individuals. Confusingly the term cluster analysis, most commonly used to describe them, has sometimes also been used for modified correlational methods concerned with associations between variables, which are quite different in concept.

There is a similar philosophy behind all of the techniques. The investigator will have measured on a set of variables the individuals he wishes to classify. Each technique contains some statistical criterion measuring similarity or separation between individuals and/or groups, and a method of evaluating each individual and forming groups so as to maximize the criterion. Essentially this is asking the data to form their own classification. It is a circular kind of operation not really susceptible to formal testing of hypotheses or statistical significance. Like factor analysis it is more in the nature of an exploratory procedure. A traditional classification may be considered to relieve support if it re-emerges from the analyses; a new one if it illuminates relationships and has clinical meaning. Validation of a new classification will involve further studies, showing that it is replicable, and that further information can be generalized from it.

Cluster analysis presents some additional problems (Everitt, 1972). It is statistically powerful and may produce chance findings. Interpretation of what the groups mean is subjective, but so is that of dimensions in factor analysis. Unlike factor analysis, it is relatively young, and a standard technique has yet to become established among the competing possibilities. Technique matters and different techniques may produce different results (Strauss *et al*, 1973). This is because the number of possible groupings of say 200 individuals measured on ten variables is immense, so all techniques employ some kind of statistical or computational shortcut.

Cluster analysis came into psychiatry comparatively recently in the late 1960s, in studies of borderline states (Grinker *et al*, 1968), depression (Paykel, 1971; Pilowsky *et al*, 1969), and psychoses (Lorr, 1966).

A third technique is discriminant analysis. Here the groups must already be decided. Their characteristics are then measured on certain variables. Discriminant analysis will produce a discriminant function (more than one, if there are more than two groups) which gives the linear combination of weighted variables which will produce the best separation between the groups. This corresponds to another question: given that we have decided on groups, how well are they separated? Are borderline cases rare, in which case an index spanning two groups will show bimodal distribution, with an area of rarity in between, or are they common, in which case a distribution will be unimodal, with frequent occurrence of intermediate cases. Any single variable or ad hoc combination of variables can be used for this purpose. Discriminant functions may not provide the ideal tool, because the

https://doi.org/10.1192/bjp.139.4.357 Published online by Cambridge University Press

analysis is very powerful and can maximize chance findings to produce good separation. There are statistical tests for the significance of separation and of non-normal distribution, but not of bimodality. Findings of bimodality on discriminant function analysis really need replication using the same score in a new sample. This has rarely been attempted and even more rarely been achieved. Discriminant function analysis is a relatively old and well established technique: it was introduced into psychiatry by Rao and Slater (1949) and into the debate about depression in the form of a related technique, multiple regression analysis by Carney *et al* (1975).

The studies

Many of the studies are well known and I will concentrate on the more recent ones. The area most studied has been depression. Among the many factor analytic studies the majority have found one or more factors corresponding to psychotic (endogenous) and neurotic (reactive) depression (Mendels and Cochrane, 1968). Studies using discriminant functions and other scores to examine bimodality of distributions have been more equivocal. The Newcastle group and its offshoots have found bimodality; others have not (Carney *et al*, 1965; Garside *et al* 1971; Andrews *et al*, 1973; Kendell, 1968; Kendell and Gourlay, 1970; Kendell and Post, 1973; Paykel *et al*, 1971; Rosenthal, 1967; Ni Brolchain *et al*, 1979).

Cluster analytic studies are fewer but growing. Pilowsky *et al* (1969) using self-report data from a mixed patient sample, obtained groups of endogenous depressives, mixed depressives and non-depressives. Paykel (1971) obtained four groups of depressives: psychotic depressives, anxious depressives, hostile depressives, young depressives with personality disorder. Overall *et al* (1966) using between-person factor analysis obtained three groups, retarded, anxious and hostile depressives. Raskin and Crook (1976) using a similar technique obtained four groups identified as endogenous, neurotic, agitated and young poor premorbid personality. Demel *et al* (1973) briefly described an analysis which produced three types of depression: endogenous, neurotic, and mildly ill. Byrne (1978), using self-report data, found four clusters: severe endogenous depressives, less severe endogenous depressives, involutional melancholics, patients with depression as symptom rather than syndrome. Andreasen *et al* (1980) obtained three groups interpreted as endogenous, neurotic and mild. The most striking feature is the consistent emergence of a group of endogenous or psychotic depressives, usually severely ill. The second feature is the general failure of non-endogenous depression to emerge as a single group rather than in several groups. There is also clear overlap in some of the additional groups which have emerged in different studies.

The discrimination between depression and anxiety has provided the second area for multivariate studies. The best known are the Newcastle Studies (Roth *et al*, 1972; Gurney *et al*, 1972). Anxiety and depressive symptoms loaded at opposite poles of a bipolar factor. Scores on this factor were not bimodally distributed but scores on a discriminant function were. A newer study with similar findings is reported by Roth and Mountjoy (in press). Mendels *et al* (1972) found self-rating scales for depression and anxiety correlated highly and separate factors did not emerge. Prusoff and Klerman (1974) found significant separation on a discriminant function but did not examine for bimodality of distribution.

The boundary between schizophrenia and affective disorders has been examined in two studies. Kendell and Gourlay (1970) failed to obtain bimodal distribution on a discriminant function based on history and mental state. Brockington *et al* (1979) did find bimodality on a similar score in a general psychotic sample but not when the score was applied to a schizoaffective sample, a rather stringent test.

Schizophrenia itself has received less study by these means, an exception being a series of studies by Lorr and colleagues (Lorr, 1966). Carpenter *et al* (1976), using data from the International Pilot Study of schizophrenia found traditional schizophrenia subtypes to be poorly separable. A cluster analysis produced strongly differentiated groups.

A number of studies have sought to classify suicide attempters. Bagley (1973) interpreted two factor analytic dimensions as indicating medically serious attempts associated with old age and depression; and attempts associated with anti-social personality and alcoholism in younger males. Fahy *et al* (1970) interpreted five factors. Katschnig and Sint (1974) derived by cluster analysis seven groups which coalesced into a large group of female overdosers with conflict motives and a smaller group of older retired persons with unhappiness motives. Kiev (1976) described seven groups. Henderson *et al* (1977) described a depressed alienated group with high risk to life; a group with low depression, low risk to life and operant motivation; and a non-specific group. In a later study (Henderson and Lance, 1979) three additional groups were identified: repeaters, wrist cutters, and an operant non-alienated group. Paykel and Rassaby (1978) found three groups: overdosers with less risk to life and interpersonal motivation; patients using violent methods, with more risk to life and self destrictive motivation; and repeaters. None of these systems has become well established, but there is general recognition of the contrast between overdosers, making mild attempts

https://doi.org/10.1192/bjp.139.4.357 Published online by Cambridge University Press

with interpersonal motivation, and the more self-destructive attempts of the severe depressive.

There have been many other factor analytic studies of different syndromes. Some have really been concerned with data condensation in rating scales, rather than classification. In the cluster analytic field the earliest psychiatric study to use a true cluster analytic technique, rather than between-person factor analysis, appears to have been of the borderline syndrome (Grinker *et al*, 1968).

A small number of studies have examined a range of patients by cluster analysis, in an attempt to validate traditional psychiatric diagnostic categories. Everitt *et al* (1971) used two different cluster analytic techniques in two psychiatric samples. All four analyses produced clusters identifiable with mania, psychotic depression, acute paranoid schizophrenia and chronic schizophrenia. Patients with depressive and other neuroses, personality disorders or alcoholism did not form clusters. Strauss *et al* (1973) applied several different methods to a predominantly schizophrenic sample. They obtained significant overlap with ICD diagnoses, but it was not very high. Williams *et al* (1976) studied a mixed sample of patients and obtained four groups which were not very easily interpretable.

The contribution

What is the contribution of these studies and the techniques which they have adopted? Clearly it is mixed and partial rather than definitive.

A very great contribution was to lift the debate on the classification of depression from personal bias and polemic, to empirical research. The reader who doubts this might compare papers from the great debates on depression in the 1920s and 30s with the research reports of the 1960s and 70s.

Secondly, the factor analytic studies in depression demonstrated convincingly that a number of different variables were indeed associated in such a way as to contrast endogenous and neurotic depression. Factor analysis has been less productive outside this realm but has contributed to the distinction between anxiety and depression. It has also been of great use as a method of condensing data from rating scales.

The contribution of cluster analysis has so far been less definitive than might have been hoped. A major contribution was moving the field of depression from a two group view, the only one that factor analysis could handle adequately, to the possibility of multiple sub-groups. The final verdict is not yet in on cluster analysis, or on the best method. It is noteworthy that the number of papers using it has slowly increased, as factor analytic papers have much fallen off.

Perhaps the least contribution has been from discriminant function analysis. The technique is so powerful statistically that the results do not always seem plausible and are hard to replicate. It may be that the question asked, as to whether there is bimodality of distribution and clear separation is not a very useful one. It is the manifest absence of clear-cut distinctions and the frequency of borderline cases which in the first place necessitated recourse to multivariate statistics. Until we get some other more fundamental measure of underlying pathology, biochemical, functional or structural, there is a case for abandoning the debate, accepting that the syndromal distinctions are not very precise, but form a useful framework. At best there are boundary zones of relative infrequency.

One thing that the techniques have not done is to firmly establish entirely new disorders. Suspected subgroups have emerged more clearly as a result of these analyses. Less suspected groups have been given some impetus towards crystallization, but have not yet reached the official classificatory schemata. The great syndromes in psychiatry were identified by penetrating clinical observers, not applied statisticians. There may yet be more symptomatic syndromes awaiting discovery, where the retentive observer's eye is likely to be more illuminating than the blanket statistical technique, which can easily miss infrequent exceptional individuals. However, most of the opportunity for such insights is past. In the tidying up that remains, multivariate statistical techniques provide a useful set of adjunctive tools for empirical research.

References

ANDREASEN, N. C., GROVE, W. M. & MAURER, R. (1980) Cluster analysis and the classification of depression. *British Journal of Psychiatry*, **137**, 256–65.

ANDREWS, G., KILOH, L. G. & NEILSON, M. (1973) Patterns of depressive illness. The compatibility of disparate points of view. *Archives of General Psychiatry*, **29**, 670–3.

BAGLEY, C. (1973) Social policy and the prevention of suicidal behaviour. *British Journal of Social Work*, **3**, 473–95.

BROCKINGTON, I. F., KENDELL, R. E., WAINWRIGHT, S., HILLIER, V. F. & WALKER, J. (1979) The distinction between the affective psychoses and schizophrenia. *British Journal of Psychiatry*, **135**, 243–8.

BYRNE, D. G. (1978) Cluster analysis applied to self-reported depressive symptomatology. *Acta Psychiatricia Scandinavica*, **57**, 1–10.

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.

CARPENTER, W. T., BARTKO, J. J., CARPENTER, C. L. & STRAUSS, J. S. (1976) Another view of schizophrenia sub-types. A report from the International pilot study of schizophrenia. *Archives of General Psychiatry*, **33**, 508–16.

https://doi.org/10.1192/bjp.139.4.357 Published online by Cambridge University Press

DEMEL, I., GORDESCH, J., KATSCHNIG, H., POLDINGER, W., SCHEIBER, V. & SINT, P. (1973) Multivariate procedures in the statistical analysis of psychic and somatic symptoms in depression. In *Psychotherapy and Psychosomatics* (eds. J. Ruesch, and A. H. S. Schmale). Karger: Basel.

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

—— GOURLAY, A. J. & KENDELL, R. E. (1971) An attempt at validation of traditional psychiatric syndromes by cluster analysis. *British Journal of Psychiatry*, **119**, 399–412.

FAHY, T. J., BROCKLEBANK, J. T. & ASHBY, D. W. (1970) Syndromes of self-poisoning. *Irish Journal of Medical Science*, **3**, 497–503.

GARSIDE, R. F. & ROTH, M. (1978) Multivariate statistical methods and problems of classification in psychiatry. *British Journal of Psychiatry*, **133**, 53–67.

—— KAY, D. W. K., WILSON, I. C., DEATON, I. D. & ROTH, M. (1971) Depressive syndromes and the classification of patients. *Psychological Medicine*, **1**, 333–8.

GRINKER, R., WERBLE, B. & BRYCE, R. E. (1968) *The Borderline Syndrome*. New York: Basic Books.

GURNEY, C., ROTH, M., GARSIDE, R. F., KERR, T. A. & SHAPIRA, K. (1972) Studies in the classification of affective disorders. The relationship between anxiety states and depressive illness, II. *British Journal of Psychiatry*, **121**, 162–6.

HAMILTON, M. (1960) A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, **23**, 56–62.

—— & WHITE, J. M. (1959) Clinical syndromes in depressive states. *Journal of Mental Science*, **105**, 985–98.

HEMPEL, C. G. (1961) Introduction to problems of taxonomy. In *Field Studies in the Mental Disorders* (ed. J. Zubin). pp 3–22. New York: Grune & Stratton.

HENDERSON, A. S. & LANCE, G. N. (1979) Types of attempted suicide (parasuicide). *Acta Psychiatrica Scandinavica*, **59**, 31–9.

—— HARTIGAN, J., DAVIDSON, J., LANCE, G. N., DUNCAN-JONES, P., KOLLER, K. M., RITCHIE, K., MCAULEY, H., WILLIAMS, C. L. & SLAGHUIS, W. (1977) A typology of parasuicide. *British Journal of Psychiatry*, **131**, 631–41.

KATSCHNIG, H. & SINT, P. (1974) Are there different types of attempted suicide? A cluster analytic approach. In *Proceedings 7th International Conference for Suicide Prevention, Amsterdam 1973* (eds. Speyer, Diekstra and Van der Loo). Amsterdam: Swets and Zeitlinger BV.

KENDELL, R. E. (1968) The classification of Depressive Ilnesses. Institute of Psychiatry, Maudsley Monograph, No. 18 (Oxford University Press).

—— (1975) *The Role of Diagnosis in Psychiatry*. Oxford: Blackwell Scientific Publications.

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

—— & POST, F. (1973) Depressive illnesses in late life. *British Journal of Psychiatry*, **122**, 615–17.

KIEV, A. (1976) Cluster analyses profiles of suicide attempters. *American Journal of Psychiatry*, **133**, 150–3.

LORR, M. (1966) *Explorations in Typing Psychotics*. Oxford: Pergamon Press.

MENDELS, J. & COCHRANE, C. (1968) The nosology of depression: the endogenous-reactive concept. *American Journal of Psychiatry*, **124**, 1–11.

MENDELS, J., WEINSTEIN, N. & COCHRANE, C. (1972) The relationship between depression and anxiety. *Archives of General Psychiatry*, **27**, 649–53.

MOORE, T. V. (1930) The empirical determination of certain syndromes underlying praecox and manic-depressive psychosis. *American Journal of Psychiatry*, **9**, 719–38.

NÍ BHROLCHAÍN, M., BROWN, G. W. & HARRIS, T. O. (1979) Psychotic and neurotic depression: 2. Clinical characteristics. *British Journal of Psychiatry*, **134**, 94–107.

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**, 946–64.

PAYKEL, E. S. (1971) Classification of depressed patients: a cluster analysis derived grouping. *British Journal of Psychiatry*, **118**, 275–88.

—— & RASSABY, E. (1978) Classification of suicide attempters by cluster analysis. *British Journal of Psychiatry*, **133**, 45–52.

—— PRUSOFF, B. & KLERMAN, G. L. (1971) The endogenous-neurotic continuum in depression: rater independence and factor distributions. *Journal of Psychiatric Research*, **8**, 73–90.

PILOWSKY, I., LEVINE, S. & BOULTON, D. M. (1969) The classification of depression by numerical taxonomy. *British Journal of Psychiatry*, **115**, 937–45.

PRUSOFF, B. & KLERMAN, G. L. (1974) Differentiating depressed from anxious neurotic out-patients. Use of discriminant function analysis for separation of neurotic affective states. *Archives of General Psychiatry*, **30**, 302–9.

RAO, C. R. & SLATER, P. (1949) Multivariate analysis applied to differences between neurotic groups. *British Journal of Psychology* (Statistical Section), **2**, 17–29.

RASKIN, A. & CROOK, T. H. (1960) The endogenous-neurotic distinction as a predictor of response to antidepressant drugs. *Psychological Medicine*, **6**, 59–70.

ROSENTHAL, S. H. (1967) Neurotic and endogenous depressions: a sceptical view. *British Journal of Psychiatry*, **113**, 1154–5.

ROTH, M. & MOUNTJOY, C. Q. (in press) The distinction between anxiety states and depressive disorders. In *Handbook of Affective Disorders* (ed. E. S. Paykel). Edinburgh: Churchill Livingstone.

https://doi.org/10.1192/bjp.139.4.357 Published online by Cambridge University Press

—— GURNEY, G., GARSIDE, R. F. & KERR, T. A. (1972) Studies in the classification of affective disorders. The relationship between anxiety states and depressive illness. *British Journal of Psychiatry*, **121**, 147–61.

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

WILLIAMS, G. W., BARTON, G. M., WHITE, A. A. & WON, H. (1976) Cluster analysis applied to symptom ratings of psychiatric patients: an evaluation of its predictive ability. *British Journal of Psychiatry*, **129**, 178–85.

WITTENBORN, J. R., HOLZBERG, J. M. & SIMON, B. (1953) Symptom correlates for descriptive diagnosis. *Genetic Psychology Monographs*, **47**, 237–301.

E. S. PAYKEL

https://doi.org/10.1192/bjp.139.4.357 Published online by Cambridge University Press