British Journal of Psychiatry (1989), **155**, 48–54

DSM–III Major Depressive Disorder in the Community
A Latent Class Analysis of Data from the NIMH Epidemiologic Catchment Area Programme

WILLIAM W. EATON, AMY DRYMAN, ANN SORENSON and ALLAN McCUTCHEON

The fit of the structure of DSM–III major depressive disorder to data from two large epidemiological surveys is assessed by latent class analysis. The surveys were conducted at the Baltimore and Raleigh–Durham sites of the National Institute of Mental Health (NIMH) Epidemiologic Catchment Area Program. Three classes are required to fit the data, and the third class bears a strong resemblance to major depressive disorder, although it requires slightly more symptoms to be present than DSM–III. The derived structure replicates successfully for Baltimore and Raleigh–Durham, with a prevalence of the major depression category of 0.9% for both sites.

Empirical evidence for the manner in which signs and symptoms cluster into purported diagnostic categories has been termed *internal construct validity* by Young (1983). Although there is considerable logic in favour of quantitative studies of the taxonomy of psychiatric disorders such as depression (e.g. as Pfohl and Andreasen argue in their 1978 paper), by 1976 the results were so ambiguous that Kendell could state that ". . . the forty or so factorial studies of depressive symptomatology that have been performed in the last twenty years leave us little the wiser" (1976, p. 21). Many of the studies reviewed by Kendell were factor analyses and cluster analyses conducted in the late 1960s and early 1970s (e.g. Eysenck, 1970; Paykel, 1971; Everitt *et al*, 1971; Fleiss, 1972). Reviewing much of the same evidence in 1982, Andreasen was slightly more positive in concluding that the studies "give some mathematical validation to at least one traditional subtype (the severe, psychotic, or endogenous)".

Several new methods of quantitative classification have been developed since these research studies and reviews were conducted, including dichotomous factor analysis (Muthen, 1978), latent trait analysis (Duncan-Jones *et al*, 1986), grade of membership analysis (Woodbury *et al*, 1978), multidimensional scaling (Schiffmann *et al*, 1981) and latent class analysis (McCutcheon, 1987). All these methods have important advantages over standard factor and cluster analysis (for an overview, see Everitt, 1984 or Bartholomew, 1987). Although we cannot present a comparison of advantages and disadvantages of these methods, which collectively we term latent structure models, we will present and discuss several important advantages of the method used here, latent class analysis.

One important weakness in almost all studies of the internal construct validity of depressive disorders is that the data on signs and symptoms were obtained from patients being treated in psychiatric settings. Although psychiatric epidemiologic surveys have been conducted for at least a century, the data obtained have been insufficient for numerical taxonomic analysis. In some cases symptom checklists have been obtained from community samples (e.g. Radloff, 1977), but the resulting data do not provide an adequate breadth of coverage of psychopathology to be informative about the specific categories of affective disorders. In some epidemiologic studies there has been adequate information to make diagnoses, but the material on signs and symptoms has not been retained to allow empirical study of the covariance of signs and symptoms.

This feature of research on depression – that empirical taxonomic studies are conducted on samples of patients – is not trivial. In deciding how to formulate a diagnostic category, the relevant base for the formulation should be the entire population, not the treated population. There are presumably important and strong selection biases for treatment for depression, as there are for most illnesses. In all probability, treated cases have more signs and symptoms than untreated cases, and the signs and symptoms they have are more severe. It also seems likely that treated cases have more psychopathology and physical illness outside the area of depression. Although these biases have not yet been conclusively established, they are so reasonable as to vitiate the value of the study of internal construct validity with samples of treated cases.

In this analysis we take advantage of data from the NIMH Epidemiologic Catchment Area (ECA) Program. These data are unique in including reports, from large community samples, on symptoms explicitly related to the diagnostic category of major depressive disorder, as laid out in DSM-III (American Psychiatric Association, 1980). Thus, our major research question here is: can we 'give some mathematical validation' to the most important and traditional subtype (major depressive disorder), with improved methods of classification and data from community samples?

The Epidemiologic Catchment Area Program

The Epidemiologic Catchment Area (ECA) Program is a series of epidemiological surveys conducted by university-based researchers in five community mental health centre catchment area populations (Eaton *et al*, 1981). At each site interviews were conducted with probability samples of about 3000 individuals living in the household population and about 500 individuals living in institutions. Each site included at least two rounds of interviews separated by one year. The research design is described in Eaton *et al* (1981) and in Eaton & Kessler (1985). Data presented below are drawn from the household sample of the Johns Hopkins ECA site in Baltimore, Maryland, and the Duke site in Raleigh-Durham, North Carolina.

The first wave of household interviewing yielded 3481 completed interviews in Baltimore for a response rate of 78%, and 3921 in Raleigh-Durham, with a 77% response rate. Initial analyses of data on non-response (Von Korff *et al*, 1985) do not reveal strong biases among sociodemographical or psychopathological variables according to non-response or attrition status. There were complete data on all relevant symptoms on 3198 subjects in Baltimore and 3614 subjects in Raleigh.

The section of the interview on psychopathology common to all the ECA sites was drawn from the NIMH Diagnostic Interview Schedule (DIS) (Robins *et al*, 1985). The DIS is a highly structured interview designed to resemble a typical psychiatric interview and to yield similar results in terms of specific mental disorder diagnoses. Results of a study of inter-rater agreement of the DIS conducted with a sample of patients in a clinical setting produced moderately good estimates of test-retest concordance for diagnoses (Robins *et al*, 1981).

Latent Class Analysis

In their classic discussion Lazarsfeld & Henry (1968) define a latent variable as an unobserved variable that explains the association among a set of observed variables. This definition is consistent with the notion of a 'factor' in factor analysis, which is estimated so as to minimise covariation among observed variables, after adjusting for the factor score. The approach in factor analysis assumes that variables are normally distributed and related to each other in a linear and additive fashion. In latent class analysis the variables are assumed to be categorical in nature, and there are no assumptions of linearity or additivity.

Observed covariation among categorical variables can be analysed by an n-way cross-tabulation, where n is the number of observed variables. In such a cross-tabulation, a standard Pearson χ^2 test of departure from independence states whether there is association among the variables. If there is association, there might exist some other variable X, with, say, T classes, such that the association in the n-way cross tabulation was zero or trivial within each class of X. If X is observable, it may be concluded that it has explained the association among the n variables, according to the standard logic of survey analysis (Rosenberg, 1968). If X is not observable, it meets the definition of the latent variable set forth by Lazarsfeld & Henry (1968).

In latent class analysis, iterative fitting algorithms such as were developed by Goodman (1974) and Clogg (1977) seek latent variables that are successful in eliminating or reducing within-class association among observed variables. Use of the likelihood ratio χ^2 yields a statistical measure of the goodness of fit. The analysis yields two sets of important parameters. One set of parameters is the proportion in the sample in the various T classes of the unobserved X variable. In the epidemiological framework these parameters are prevalences. Another set of parameters is the set of probabilities that the observed variables take values, given that the individual is a member of an unobserved class. Clogg (1979) has suggested that these conditional probabilities are similar to the concept of factor loadings in factor analytic models.

There is a confirmatory aspect to latent class analysis. By requiring that certain conditional probabilities equal unity or zero (referred to as deterministic constraints), explicit relationships can be incorporated into the analysis. Conditional probabilities in different classes can also be constrained to equal one another (equality constraints). Thus, a given diagnostic model can be tested in a confirmatory mode. The question of interest becomes: if the sample is divided into those subjects meeting criteria for diagnosis and those not meeting criteria, is the association of the symptom variables reduced to a non-significant level?

Latent class analysis has certain important advantages over traditional factor analysis. It does not require or assume that the data are normally distributed, which is rarely or ever true in the case of psychopathology. Moreover, the measures of the association used are based on the odds-ratio, which means that they are not influenced by the prevalence of the characteristic under study. Thus, taxonomic results can be compared for samples which differ in their prevalence (e.g. samples of men and women; samples of subjects under treatment against those in the community). Finally, there is no assumption of additivity in the association between variables. Since many diagnostic categories may not necessarily conform to dimensional concepts, this advantage of latent class analysis is a crucial one.

Results

First the prevalence of individual symptoms and DSM-III symptom groups is described. In the DIS, the phrasing of all symptom questions is in the lifetime format, beginning with the phrase "Have you ever? . . ." If the answer is positive, questions follow to ensure that the symptom meets criteria of severity, and that it was not caused by drugs, medication, alcohol, physical illness, or injury. If these conditions are met, the symptom is considered to be a "plausible psychiatric symptom" (Robins *et al*, 1985). After all symptoms in the general area of depression are enquired about, the respondent is asked whether the symptoms ever occurred together, and if so, whether they were dysphoric during that particular time. Then, there are questions on the first and most recent occurrence of 'these problems'. The symptom data can be used to analyse the lifetime prevalence of DSM-III major depressive disorder (Robins *et al*, 1984).

There are anomalies in the analysis of lifetime prevalence data that lead us to be very cautious about their interpretation and analysis. For example, the lifetime prevalence of major depressive disorder declines with increase in age (Robins *et al*, 1984), which indicates either a high mortality rate, or a large cohort effect, or a tendency for older persons to forget or deny symptoms which have occurred in the distant past. These problems led us to focus on the symptoms occurring within the month prior to the interview. This was made possible by the implementation of probes for recall of symptom occurrence used in several of the ECA sites (Von Korff & Anthony, 1982). Since this information is available for every symptom, there is little possibility of error due to failure to recall the timing and clustering of the individual symptoms in a given episode.

Table I presents data on the prevalence of the DIS symptoms relevant to major depressive disorder and the prevalence of the DSM-III symptom groups for all respondents in Baltimore. The table presents abbreviated descriptions of the DIS questions – most of them include several phrases relevant to the symptom in question, and almost all include the phrase "a period of two weeks or more when you. . .". For example, DIS Question 72, which is the indicator for DSM-III criterion A, is worded as follows: "In your lifetime, have you ever had two weeks or more in which you felt sad, blue, depressed, or when you lost all interest and pleasure in things that you usually cared about or enjoyed?" This symptom is referred to in the table 'dysphoria'. As described above, the symptoms are recorded as having been present in the month prior to the interview, having met the DIS criteria of severity, and, where appropriate, causes of the symptom due to medication, drugs, alcohol, physical illness, or injury have been ruled out according to the DIS method (Robins *et al*, 1985). The group prevalences record the presence of any symptom in the group, according to the same criteria, so that these prevalences are constrained to be equal to or greater than the symptom prevalences.

TABLE I
Prevalence of dysphoria and eight DIS/DSM-III symptom groups for all respondents in the Epidemiologic Catchment Area of East Baltimore

Symptom or group title, DIS no. and content of DIS question	*Percentage prevalence of symptom*	*Percentage prevalence of any symptom in the group*
Dysphoria		4
72 Sad for 2 weeks	4	
Group 1		6
74 Loss of appetite	2	
75 Lose 2 lbs/week	1	
76 Gain 2 lbs/week	4	
Group 2		11
77 Trouble falling asleep	9	
78 Sleeping too much	2	
Group 3		6
80 Talk/move slowly	3	
81 Move all the time	3	
Group 4		2
82 Loss of interest in sex	2	
Group 5		7
79 Tired out	7	
Group 6		3
83 Worthless	3	
Group 7		5
84 Trouble concentrating	4	
85 Thoughts slow	3	
Group 8		9
86 Thought about death	8	
87 Wanted to die	3	
88 Thought of suicide	1	
89 Attempted suicide	0	

The most frequent symptoms are 'trouble falling asleep' (9%), 'thought about death' (8%), and 'tired out' (7%). Only one symptom, 'attempted suicide', had a prevalence which rounded to zero. In some symptom groups the overlap of symptoms is already apparent: for example, in groups 7 and 8, the prevalence of the group is only slightly higher than the most prevalent symptom. In other groups the substitutability of the symptoms is apparent: for example, in groups 2 and 3, respondents tend to have one or the other symptom, but not both, so that the group prevalence is the sum of the symptom prevalences.

TABLE II

Dysphoria and eight DIS/DSM-III symptom groups: most frequent combinations of data for all respondents in the Epidemiologic Catchment Area of East Baltimore

Cell number	*Frequency*	*Cumulative percentage*	*DIS/DSM-III symptom groups in which symptoms occurred*[1]
1	2393	74.8	No symptoms
129	122	78.7	2
3	107	82.0	8
17	52	83.6	5
257	50	85.2	1
65	29	86.1	3 8
5	30	87.0	7
131	21	87.7	2 8
385	16	88.2	1 2
145	15	88.7	2 5
2	14	89.1	D
33	14	89.6	4
67	11	89.9	3 8
81	10	90.2	3 5

1. D dysphoria; 1 appetite loss or weight change; 2 insomnia or hypersomnia; 3 psychomotor agitation or retardation; 4 loss of interest; 5 fatigue; 6 feelings of worthlessness; 7 concentration problems; 8 suicidal ideation.

How might these symptoms be grouped in the population? With eight groups and dysphoria, there are two to the ninth power or 512 possible patterns of responses. Table II shows that, next to no symptom groups at all (cell one with 75% of the sample), the most frequent response patterns are the occurrence of a single symptom group. Seven patterns (no symptoms; group 2 only; group 8 only; group 5 only; group 1 only; group 3 only; and group 7 only) account for 87% of the sample. Another 7% are accounted for by five patterns showing the occurrence of two symptom groups only, and two more patterns with one symptom only. Thus, out of 512 possibilities, 14 patterns account for 90% of the sample.

Of the 512 possible patterns, 256 do not include dysphoria and will, therefore, not be eligible for meeting DSM-III criteria for a major depressive episode. Of the 256 patterns which include dysphoria, combinatorial logic shows that 163 meet DSM-III criteria and 93 do not. The figure 163 is the sum of the possible combinations of eight things taken eight at a time (1); seven at a time (8); six at a time (28); five at a time (56); and four at a time (70). How frequently do these patterns occur in the sample?

Table III shows all combinations of symptom groups meeting DSM-III criteria for major depressive disorder which actually occurred in the sample, in order of frequency of occurrence. Of the 163 possible combinations, only 39 actually occurred in the sample. There were 49 cases in this sample, for a prevalence of DIS/DSM-III depressive disorder of 1.5%. Several symptom groups are relatively more common among these 49 DSM-III cases, including groups 2 (sleep problems – 41 cases) and 7 (concentration – 36 cases). Two groups are much less common: loss of pleasure, present in 10 cases; and feelings of worthlessness, present in 24 of the 49 cases.

TABLE II

Dysphoria and eight DIS/DSM-III symptom groups: combinations of data meeting criteria for DSM-III major depressive disorder, in order by frequency for all respondents in the Epidemiologic Catchment Area of East Baltimore

Cell number	*Frequency*	*DIS/DSM-III symptom groups in which symptoms occurred*							
480*	3	1	2	3		5	6	7	8
352*	3	1		3		5	6	7	8
454	3	1	2	3				7	
472*	2	1	2	3		5		7	8
408*	2	1	2			5		7	8
256*	2		2	3	4	5	6	7	8
212	2		2	3		5			8
512*	1	1	2	3	4	5	6	7	8
502*	1	1	2	3	4	5		7	
498*	1	1	2	3	4	5			
496*	1	1	2	3	4		6	7	8
478*	1	1	2	3		5	6	7	
468*	1	1	2	3		5			8
462*	1	1	2	3			6	7	
456*	1	1	2	3				7	8
416*	1	1	2			5	6	7	8
400*	1	1	2				6	7	8
398*	1	1	2				6	7	
350*	1	1		3		5	6	7	
348*	1	1		3		5	6		8
248*	1		2	3	4	5		7	8
224*	1		2	3		5	6	7	8
220*	1		2	3		5	6		8
216*	1		2	3		5		7	8
214*	1		2	3		5		7	
200*	1		2	3				7	
160*	1		2			5	6	7	8
144*	1		2				6	7	8
84*	1			3		5	6	7	
88*	1			3		5		7	8
482	1	1	2	3	4				
466	1	1	2	3		5			
452	1	1	2	3					8
396	1	1	2				6		8
406	1	1	2			5		7	
242	1		2	3	4	5			
180	1		2		4	5			8
152	1		2			5		7	8
92	1			3		5	6		8
Total number of									
cells	39	23	33	29	9	28	19	27	26
cases	49	31	41	38	10	36	24	36	34

*Cell assigned to class C of restricted (C) model in Table IV.

The cross-tabulation of every symptom group (including dysphoria) with every other group produced a 512-cell table which was submitted to the Maximum Likelihood Latent

TABLE IV
Latent class analysis of model fitting results for data on depression for the Epidemiologic Catchment Area of East Baltimore

Model	χ^2	*Degrees of freedom*	*Class prevalence* A	B	C	D
Independence	2416	511				
Two classes						
Unrestricted	559	492	88	12	–	–
Restricted (DSM–III)	1045	493	97	3	–	–
Three classes						
Unrestricted	400	482	83	15	2	–
Restricted (C)	436	483	86	12	1	–
Restricted (BC)	995	485	97	2	1	–
Four classes						
Unrestricted	376	474	82	14	2	2
Restricted (C)	385	473	82	15	1	2

Structure Analysis Program (Clogg, 1977) as described above. (The programme had to be re-dimensioned to accept so large a table.) Table IV reports the results of several latent class analyses of these data.

The model of independence produces a large χ^2 value of 2416, indicating that there is a substantial amount of covariation among the symptoms groups. If we require the algorithm to fit a model with just two classes, a large amount of that covariation is explained – χ^2 for this unrestricted model is 559, with 492 degrees of freedom. This fit is surprisingly effective but nevertheless unacceptable because χ^2 is too large. Examining the conditional probabilities in this model (not shown) reveals class A to be the 'normal' population, since the probabilities are so low for those in their class. Class B is the 'depressives', comprising 12% for the sample. If we place a restriction in the model to require that dysphoria be present in the class of 'depressives', in effect mimicking the structure of DSM–III, χ^2 is increased to 1045, even though the prevalence of 3% is in accord with common wisdom. If the restricted two-class model fits the data well (e.g. with χ^2 below 500), that would mean that all the covariation in these symptom groups could be explained by the simple presence or absence of a single disorder – a strict, simple, and elegant model.

An unrestricted three-class model fits the data fairly well, with a χ^2 value of 400. The class of 'normals' has a prevalence of 83%, and there is a third class (C) which includes severely depressed individuals with high conditional probabilities of symptom occurrence, at a prevalence of 2%. The intermediate class B has a prevalence of 15%. The second three-class model is a restricted model and is labelled 'C' because dysphoria is constrained to be present in class C. This model represents the closest approximation to the diagnostic structure in DSM–III and it is therefore discussed in more detail after an examination of the fit of the single remaining three-class model and certain four-class models. A final three-class model labelled 'restricted – BC' constrains dysphoria to be present in both of the depressive classes. The fit is unacceptable with a χ^2 value of 995.

Two four-class models were tested and they fit the data acceptably. The unrestricted four-class model produced a χ^2 value of 376. There are two 'severe' classes in this model with prevalence of 2% each, and an intermediate class with prevalence of 14%. A four-class model with dysphoria constrained to be present in the third class also fits the data well. But in neither of these models was there an important improvement in fit over the three-class models, and since they are less parsimonious, we focus on the three-class restricted model.

The value of χ^2 in this circumstance is somewhat compromised as a test of fit because of the large number of cells in the table with expected frequencies less than five. The impact is to overestimate the effective degrees of freedom, and be less likely to reject the fit. Where models are hierarchical, and χ^2 values with degrees of freedom are obtained by subtraction, this problem is less threatening. Therefore we are somewhat cautious about the overall fit, but impressed with the significant improvement in fit of the three-class restricted (C) model over the three-class unrestricted model. These two models are hierarchical, so that the appropriate value of χ^2 is 36 with one degree of freedom.

Table V presents the probabilities for positive responses, conditional on class membership, associated with the three-class model of interest (restricted – C). The probability of 1.0 for dysphoria signifies the restriction that it always be present, consistent with DSM–III structure. Other probabilities are also quite high, with the exception of loss of interest in sex. Class C is a slightly more conservative definition of depression than the DSM–III, in that every single response pattern which belongs within class C is consistent with the definition of DSM–III. This inclusion is denoted by an asterisk by the cell number in Table III. All response patterns with dysphoria and *five* or more symptom groups are included in class C (and in DSM–III major depressive disorder). But certain response patterns with dysphoria and *four* symptom groups are in DSM–III and not in class C (patterns 454, 212, 482, 466, 452,

TABLE V
Latent class analysis for a three-class model with dysphoria constrained to be present in the third class: conditional probabilities for positive responses for the Epidemiologic Catchment Area of East Baltimore

DSM–III group	*Class* I	II	III
Dysphoria	0.004	0.192	1.000
Group 1 – Weight	0.019	0.283	0.648
Group 2 – Sleep	0.045	0.475	0.779
Group 3 – Slow	0.009	0.314	0.778
Group 4 – Sex	0.005	0.129	0.171
Group 5 – Tired	0.017	0.379	0.764
Group 6 – Worthless	0.002	0.153	0.603
Group 7 – Concentration	0.010	0.256	0.873
Group 8 – Death	0.041	0.357	0.744

396, 406, 242, 180, 152, 92). All but one of these patterns include group 2 (problem sleeping). This suggests that problems in sleeping may be an indicator of a less serious form of disorder – a disorder which does not necessarily include dysphoria. There are also certain response patterns with less than five symptom groups which are potent enough to place them in class C. These include patterns 398, 214, 144, 94, and 88. All these patterns include group 7 (concentration). This suggests that concentration problems may be an indicator of a more serious form of depressive disorder.

Class B includes a variety of response patterns from the same domain of content. The most probable symptom is group 2 (problems with sleep), but thoughts about death, feeling tired, and motor activity are all also fairly probable. Dysphoria is only present in about 20% of the cases.

The method of latent class analysis allows strict comparison of taxonomic structures between different groups in the population. Here, to assess the robustness of the findings in Table V, a comparison has been made with data from the Raleigh-Durham site of the ECA Program. A model was fitted to the data from both sites simultaneously, in order to examine how similar the structures were. The conditional probabilities in the model for the two sites were so similar that models were fitted to the data requiring that conditional probabilities be equal in the two sites – this equality is called 'structural homogeneity' in the literature on latent class analysis. Thus, for example, the conditional probability of a positive response on DSM–III group 1 (weight–appetite) for those in the third (depression) class was 0.5859 in Baltimore and 0.5859 in Raleigh-Durham; for group 2 (sleep problems) both probabilities were equal to 0.7485 – both quite close to the probabilities shown in Table V for Baltimore only. It was possible to fit a model successfully with only one parameter not constrained to equality in the depression class. The exceptional parameter was for group 6 (feelings of worthlessness), which had a value of 0.6243 in the Baltimore sample and 0.2876 in the Raleigh-Durham sample. In effect, depressives in Baltimore are more likely to feel worthless, sinful, or guilty than those in Raleigh-Durham. It may be that a more useful classification for depression would include this parameter as part of the concept of dysphoria, and not as a separate symptom group.

There were also four parameters in the second class which had to be free to vary between the sites (groups 1, 3, 4, and 7). The estimated values for these parameters, while different, were not widely divergent (e.g. 0.2910 v. 0.1673 for group 1; 0.3135 v. 0.2218 for group 3). Finally, there was one parameter in class I which was allowed to differ – thoughts of death had a probability of 0.0411 in Baltimore for the 'normal' group, and 0.0238 for the 'normal' class in Raleigh-Durham. In all, six of the 27 parameters were unconstrained in the three-class model. Only one of the unconstrained parameters was in the third class.

The prevalences for the classes were allowed to vary, even as the hypothesis of structural homogeneity was being tested. But the prevalence of the third class was 0.0092 in Baltimore and 0.0090 in Raleigh-Durham – or very close to 1% prevalence in both sites. This prevalence is lower than the 3–4% reported in other analyses of ECA data (Myers *et al*, 1984) because we are requiring all symptoms to be present in the month before the interview, and using a slightly stricter definition of depression than the DIS/DSM–III method.

The likelihood ratio χ^2 for the model fitted for data from Baltimore and Raleigh-Durham together was 794.7, with 986 degrees of freedom. It is wise to be cautious in accepting a fit to data as sparse as these are, with many cells having expected frequencies less than unity. But this is a relatively large sample, and even with 200 less degrees of freedom, the observed values and those expected under the model would not differ statistically at the level of 0.05.

Discussion

The experience which generated the DSM–III originated in clinical settings, non population-based samples. Indeed, data from household populations have never been collected in a form which allows such a direct test of a diagnostic model. Class C above is quite close to DSM–III major depressive disorder, although perhaps slightly more severe. This amounts to a statistical and community-based confirmation of the general concept of depressive disorder, if not of exactly the DSM–III algorithm. The robustness of the findings in the comparison with the separate ECA site at Raleigh-Durham is credible, and the near equality of class prevalences is also an important result.

Class B above is some sort of intermediate disorder which does not necessarily involve dysphoria – perhaps the atypical affective disorder labelled type V by Davidson *et al* (1982). We intend to investigate this possibility and others such as dysthymia (APA, 1987), and 'neurotic' depression (Klerman *et al*, 1979) in future analyses.

Certain problems are apparent in the data and the analysis. Loss of interest in sex may be a poor indicator of loss of interest in things usually enjoyed, but it is all that is available in the data, separate from the question on dysphoria. It is not a useful symptom in the data presented above. In the revision to the DSM–III (American Psychiatric Association, 1987), the importance of sexual drive, as an indicator of more generally diminished interest or pleasure, has been de-emphasised. As with our analyses in general, we are constrained to the DSM–III symptoms as operationally defined and measured by the DIS. If the sample were large enough, and the latent structure algorithm robust enough, it would be better to study the covariation in symptoms themselves, instead of coding them into DSM–III symptom groups. But the limits of the data and the algorithm oblige us to use this group structure in the analyses. Finally, we have constrained the analyses to the single domain of depressive symptomatology included in the DIS depression section. Study of the overlap of depression with symptoms of somatisation disorder or anxiety disorders, for example, has not been attempted in this analysis. Other future research of interest includes analyses comparing population subgroups, for example according to gender and age.

This analysis reflects on the internal construct validity of the DSM-III diagnosis of major depressive disorder as measured by the DIS. Other important criteria of validity exist, of course, and all these must be weighted in any ultimate judgement of the utility of this diagnostic category.

Acknowledgements

The ECA Program is a series of five epidemiological research studies performed by independent research teams in collaboration with staff of the Division of Biometry and Epidemiology of the NIMH. During the period of data collection the ECA Program was supported by Co-operative Agreements. Dr Eaton and Dr Dryman were supported by NIMH grant number MH 41908 while this paper was being prepared.

References

AMERICAN PSYCHIATRIC ASSOCIATION (1980) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn) (DSM-III). Washington DC: APA.

—— (1987) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM-III-R). Washington DC: APA.

ANDREASEN, N. C. (1982) Concepts, diagnosis, and classification. In *Handbook of Affective Disorders* (ed. E. S. Paykel). New York: The Guilford Press.

BARTHOLOMEW, D. J. (1987) *Latent Variable Models and Factor Analysis*. London: Oxford University Press.

CLOGG, C. C. (1977) *Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users*. Pennsylvania: State College, Pennsylvania State University.

—— (1979) Some latent structure models for the analysis of likert-type data. *Social Science Research*, **8**, 287-301.

DAVIDSON, J. R., MILLER, R. D., TURNBULL, C. D., *et al* (1982) Atypical depression. *Archives of General Psychiatry*, **39**, 527-534.

DUNCAN-JONES, P., GRAYSON, D. A. & MORAN, P. A. (1986) The utility of latent trait models in psychiatric epidemiology. *Psychological Medicine*, **16**, 391-405.

EATON, W. W., REGIER, D. A., LOCKE, B. Z., *et al* (1981) The NIMH Epidemiologic Catchment Area Program. *Public Health Reports*, **96**, 319-325.

—— & KESSLER, L. G. (eds) (1985) *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program*. New York: Academic Press.

EVERITT, B. S. (1984) *An Introduction to Latent Class Models*. London: Chapman & Hall.

——, 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.

EYSENCK, H. J. (1970) The classification of depressive illness. *British Journal of Psychiatry*, **117**, 241-250.

FLEISS, J. L. (1972) Classification of the depressive disorders by numerical typology. *Journal of Psychiatry Research*, **9**, 141-153.

GOODMAN, L. A. (1974) The analysis of systems of qualitative variables when some of the variables are unobservable: Part I – a modified latent structure approach. *The American Journal of Sociology*, **79**, 1179-1259.

KENDELL, R. E. (1976) The classification of depression: a review of contemporary confusion. *British Journal of Psychiatry*, **129**, 15-28.

KLERMAN, G. L., ENDICOTT, J., SPITZER, R., *et al* (1979) Neurotic depressions: systematic analysis of multiple criteria and meanings. *American Journal of Psychiatry*, **136**, 57-61.

LAZARSFELD, P. R. & HENRY, N. W. (1968) *Latent Structure Analysis*. Boston: Houghton-Mifflin.

MCCUTCHEON (1987) *Latent Class Analysis*. Beverly Hills: Sage.

MUTHEN, B. (1978) Contributions to factor analysis of dichotomous variables. *Psychometrics*, **43**, 551-560.

MYERS, J. K., WEISSMAN, M. M., TISCHLER, G. L., *et al* (1984) Six month prevalence of psychiatric disorders in three communities. *Archives of General Psychiatry*, **41**, 959-970.

PAYKEL, E. S. (1971) Classification of depressed patients: cluster analysis derived grouping. *British Journal of Psychiatry*, **18**, 275-288.

PROHL, B. & ANDREASEN, N. C. (1978) Development of classification systems in psychiatry. *Comprehensive Psychiatry*, **19**, 197-207.

RADLOFF, L. S. (1977) The CES-D scale: a self-report depressive scale for research in the general population. *Journal of Applied Psychological Measurement*, **1**, 385-401.

ROBINS, L. N., HELZER, J. E., WEISSMAN, M., *et al* (1984) Lifetime prevalence of specific psychiatric disorders in three sites. *Archives of General Psychiatry*, **41**, 949-958.

ROBINS, L. N., HELZER, J. E., CROUGHAN, J., *et al* (1981) National Institute of Mental Health Diagnostic Interview Schedule. *Archives of General Psychiatry*, **38**, 381-389.

ROBINS, L. N., HELZER, J. E., ORVASCHEL, H., *et al* (1985) The Diagnostic Interview Schedule. In *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program* (eds W. W. Eaton & L. G. Kessler). New York: Academic Press.

ROSENBERG, M. (1968) *The Logic of Survey Analysis*. New York: Basic Books.

SCHIFFMAN, S. S., REYNOLDS, M. L. & YOUNG, F. W. (1981) *Introduction to multidimensional scaling: theory, method, and applications*. New York: Academic Press.

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**, 531-540.

VON KORFF, M. R. & ANTHONY, J. C. (1982) The NIMH Diagnostic Interview Schedule modified to record current mental status. *Journal of Affective Disorders*, **4**, 365-371.

——, COTTLER, L., GEORGE, L. K. *et al* (1985) Non-response and non-response bias in the ECA surveys. In *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program*. New York: Academic Press.

WOODBURY, M. A., CLIVE, J. & GARSON, A. (1978) Mathematical typology – a grade of membership technique for obtaining disease definition. *Computers and Biomedical Research*, **11**, 277-298.

YOUNG, M. A. (1983) Evaluating diagnostic criteria: a latent class paragon. *Journal of Psychiatric Research*, **17**, 285-296.

*William W. Eaton, PhD, *Department of Mental Hygiene, School of Hygiene and Public Health, The Johns Hopkins University*; Amy Dryman, ScD, *Department of Mental Hygiene, School of Hygiene and Public Health, The Johns Hopkins University*; Ann Sorenson, PhD, *Department of Sociology, University of Indiana*; Allan McCutcheon, PhD, *Department of Sociology, University of Delaware*

*Correspondence: *Department of Mental Hygiene, School of Hygiene and Public Health, The Johns Hopkins University, 8th Floor, Hampton House, 624 North Broadway, Baltimore, Maryland 21205, USA*