British Journal of Psychiatry (1990), **157**, 55–65

Classifying Depression by Mental State Signs

GORDON PARKER, DUSAN HADZI-PAVLOVIC, PHILIP BOYCE, KAY WILHELM, HENRY BRODATY, PHILIP MITCHELL, IAN HICKIE and KERRIE EYERS

The possibility that separation of a categorical depressive disease ('melancholia') from remaining depressive disorders can be improved by assessment of mental state signs was examined in patients treated by representative Sydney psychiatrists and patients referred to a specialised mood disorders unit. A set of signs, principally assessing retardation, was derived within the two samples by principal-components and latent-class analyses. Scores were significantly correlated with clinical, DSM–III, and RDC diagnoses, and appeared independent of severity, suggesting that melancholia can be defined phenomenologically. Scores were also associated with several 'validating' factors. Comparative analyses of a refined list of melancholia symptoms suggested that ratings of defined signs are likely to have greater capacity than symptom ratings to differentiate melancholia from residual depressive disorders.

> "Endogenous depression, perhaps renamed, appears to be a useful concept for both clinical and research purposes. It should be distinguished from non-endogenous depression". (Andreasen, 1980)

In classifying depressive disorders, a continuing objective has been to define the clinical features of an assumed categorical entity with imputed genetic and biochemical determinants (Kiloh *et al*, 1972), varyingly termed endogenous, melancholic, psychotic, severe, vital, autonomous and endogenomorphic depression, and which responds better to physical treatment. Historical accounts (e.g. Kendell, 1968; Klerman, 1971; Andreasen, 1980) of relevant theorising and research demonstrate prolonged interest in defining such an entity but, as yet, no 'gold-standard' diagnostic system has been derived.

Rosenthal & Klerman (1966) suggested that the concept of 'endogenous' or 'autonomous' depression, as proposed by Gillespie (1929), was supported by a number of factor-analytic studies and that it had come to imply: (a) a relatively stable and non-neurotic pre-morbid personality, (b) a lack of an initial precipitant, and (c) a particular pattern of symptoms and signs – lack of reactivity to environmental changes, severity of depression, feelings of guilt, remorse and feelings of unworthiness (as opposed to self-pity and externalised blame), insomnia, especially middle- and early-morning awakening, depressed mood being worse in the morning, retarded speech, thought and motor activity, motor restlessness and agitation, vegetative symptoms (e.g. constipation, dry mouth, slowed hair and nail growth), decreased libido and amenorrhoea, weight loss, and loss of interest and general lassitude. We note here the emphasis on symptoms, as against signs, in defining the construct.

While a review by Mendels & Cochrane (1968) was generally supportive of such historically promoted features of endogenous depression, or what we term here 'melancholia', a later review of factor-analysis and cluster-analysis studies by Nelson & Charney (1981) suggested a more restricted set. They held that the clearest features were psychomotor change (retardation or agitation), severity of mood, lack of reactivity, delusions, self-reproach, and loss of interest. There was some support for distinct quality, concentration difficulties, and early-morning wakening, but little support for the so-called 'vegetative' features of appetite loss, weight loss, and insomnia. Such studies have, at best, defined a dimension of variation (i.e. 'endogeneity'), but the boundary between categories of 'melancholia' and 'non-melancholic depressive disorders' has not been established with acceptable precision. Thus, when those clinically diagnosed as having 'endogenous depression' or melancholia are rated on official classificatory systems (e.g. DSM–III; American Psychiatric Association, 1980) and on discriminating rating scales, such as the Newcastle (Carney *et al*, 1965) and Michigan discriminant (Feinberg & Carroll, 1982) indices, concordance between systems tends to be moderate at best (e.g. Davidson *et al*, 1984; Zimmerman *et al*, 1986), suggesting little agreement about definition of the entity.

It is of interest that retardation, essentially a mental state item, emerged as highly discriminating in both reviews, despite few definitions. Hamilton (1967), for one, suggested that retardation could range from a "slight flattening of affect and fixity of expression", through "a monotonous voice, a delay in answering questions, a tendency to sit motionless", to a state where "retardation makes the interview extremely prolonged and almost

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

impossible''. Kiloh & Garside (1963) defined retardation as a ''subjective experience of slowness of thought or action and objective psychomotor slowing.'' While the previous focus has been on symptoms, the suggested central relevance of retardation encouraged us to determine whether typological decisions could be improved by focusing on mental state signs.

In focusing on signs rather than symptoms, some definition is in order. Kendell (1983) stated that ''most psychiatric disorders . . . are classified largely on the basis of their clinical symptoms, or to be precise, a combination of the symptoms the patient complains of and the behavioural abnormalities evident on examination''. We differentiate between symptoms and signs as is usual in general medicine, where a 'symptom' is a ''phenomenon of physical or mental disorder or disturbance which leads to complaints on the part of the patient . . . usually a subjective state'', and therefore in ''contrast to an objective sign'', where a 'sign' is ''objective evidence or physical manifestation of disease'' (*Blakiston's Pocket Medical Dictionary*, 3rd edn). Thus, to us, depressive symptoms are (generally) volunteered by the patient as self-reports, whereas depressive signs are manifestations observed by an outsider.

While there are limitations to rating signs (see later), several limitations to rating symptoms (the usual strategy) should be noted. First, distinctly different information may be provided by assessing the presence or absence of symptoms as against their severity. Second, if symptoms are rated at the nadir (which may be well in the past), the risk of retrospective falsification is considerable. Third, conversely, rating of current symptoms may be misleading if features vary over the episode. Fourth, the capacity of patients (particularly elderly or psychotic depressives) to generate valid information about symptoms may be limited by the cognitive effects of depression.

Our decision to base a typology on clinical variables was influenced in part by the views of Ní Bhrolcháin (1979), who suggested that intrinsic ambiguities in ''mixed-variable analyses can be avoided very simply by restricting the basis of classification, in the first instance, to mental state items.''

Although our final strategy is original in focusing on mental state features, a number of typology studies have included signs in their sets of clinical features. Thus, Hamilton & White (1959) assessed features such as retardation and agitation (preserved in the Hamilton Rating Scale for Depression), Rosenthal & Klerman (1966) included an unspecified number of signs in their data set, Kendell & Gourlay (1970) included in their study several behavioural ratings (e.g. monotonous voice or expressionless face, retardation of movement) from the Present State Examination (PSE) while the PSE itself (Wing *et al*, 1974) has items rating behaviour, affect, and speech.

As the history of typology research suggests the importance of replication, we impose two criteria in refining our signs: consistency across two distinct samples and across two differing multivariate techniques – principal-components and latent-class analyses. We develop a refined measure of 'melancholia' symptoms within one of our samples and then examine the extent to which our identified signs and symptoms measure a similar construct and their likely comparative utility.

A key difficulty in such inquiries is that there is no established diagnostic criterion for comparison, and yet investigation cannot proceed entirely in a vacuum. We therefore make comparisons with clinical diagnosis (and background variables). While 'psychotic' (or delusional) and 'endogenous' depressive disorders are unlikely to be identical, they are generally regarded as subtypes of melancholia and we therefore aggregate these two subgroups in many analyses.

Method

Design of the mental state rating scale

The objective was to generate signs suggestive of melancholia or endogenous despression that might be accurately judged by a clinician at interview. The items assessed: general attitude; severity of mood; responsiveness to the interviewer; retardation and agitation as expressed facially, in posture, speech, and in movement; reactivity; speech; concentration, and insight. Two features (guilt and nihilism), often regarded as symptoms rather than signs, were included with the intent that they be rated as evidenced verbally and non-verbally by the patient at interview. Because some clinicians consider that melancholics have a characteristic facies or mien, we included one item rating 'endogenous' quality (without any further description). Each item in our scale had four anchor points, with one pole assessing no disturbance and the remainder quantifying increasing levels of disturbance. Each assessed a specific behaviour rather than a behavioural domain, to reduce ambiguity and other rating problems.

Subjects

Sample A comprised 300 non-consecutive depressed patients, each assessed once by some 30 clinician raters in differing in-patient and out-patient psychiatric facilities in Sydney, and with patients also returning self-report data on depression and anxiety symptoms occurring at the nadir

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

of their episode. Additionally, patients were asked to rate the relevance of a number of life events before the onset of their depression, and to complete the Zung (1965) state depression scale. We allowed the raters a choice of only five 'clinical' diagnoses (psychotic, endogenous, neurotic, and reactive depression, and depression superimposed on anxiety), and required that they generate DSM-III and ICD-9 diagnoses (World Health Organization, 1978), with each rater being provided with listed criteria for all three diagnostic systems. In our analyses, we omitted 20 subjects who scored less than 40 on the Zung scale and who might not have been significantly depressed, and 18 subjects who returned incomplete Zung forms, resulting in an effective sample of 262 subjects.

Sample B was derived from 250 consecutively diagnosed depressive patients referred as in-patients or out-patients to our specialised mood disorders unit (MDU). Chronic and treatment-resistant depressive disorders and patients with significant personality disturbance (Brodaty *et al*, 1987) are over-represented at the MDU. Consultant psychiatrists to the MDU undertake a semistructured clinical interview, and data on depressive features are coded so that DSM-III, ICD-9, Research Diagnostic Criteria (RDC; Spitzer *et al*, 1978), and clinical diagnoses are generated, as well as scores on the Newcastle scale (Carney *et al*, 1965). The clinical diagnoses ('psychotic', 'endogenous', 'neurotic' and 'reactive' depression in this sample) are derived consensually by the MDU consultants, and are often refined to incorporate in-patient assessment, longitudinal data, and corroborative witness information. Patients complete the Zung scale, the 30-item General Health Questionnaire (GHQ; Goldberg, 1972) of psychological distress, and are rated by a psychiatrist on the 21-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1967). As a number of referred depressives are in remission at intake assessment, we elected to exclude from our analyses those 46 who scored less than 40 on the Zung scale or who failed to complete the Zung measure. Two subjects were omitted for other reasons, giving an effective sample of 202 subjects.

The clinical diagnoses for our samples are provided in Table III. Sample A subjects had a mean age of 48.3 years, a mean illness duration of 29 weeks, a mean Zung score of 56.8, and 62% were female. Sample B had a mean age of 43.5 years, a mean illness duration of 81 weeks, a mean Zung score of 55.0, and 69% were female.

Statistical analyses

In refining a set of signs we undertook two approaches: principal-components analysis (PCA), a well known strategy, and latent-class analysis (LCA), which has only recently been applied in psychiatric research. The first is useful in identifying sets of items that group together, but cannot directly address groupings of patients. The second has the potential to identify indicators separating members and non-members of a presumed latent or underlying class – in this instance, to separate those belonging to what we term 'melancholia' from residual or non-melancholic depressive disorders. While a number of LCA methods have been proposed (Goodman, 1974; Formann, 1982), we used the taxometric model of Golden (1982), judging it attractive as it was "explicitly intended for the . . . situation where one wishes to test a conjectured latent taxon of a certain substantive nature (as opposed to the situation where one searches for underlying 'clusters' of any kind)". It has also been used recently in an attempt to define endogenous depression by numerical taxonomy (Grove *et al*, 1987). Golden's procedure requires that the likely indicators of the latent category be dichotomous and positively correlated, and is undertaken iteratively. At each stage, indicators (e.g. mental state signs) that do not meet various criteria are deleted until a satisfactory set is reached. The underlying base rate of the class or category, as well as the true- and false-positive rates for each indicator, can be estimated.

From the subset of items identified by these two approaches we sought evidence of discontinuity in summed item scores, which might support the binary view of distinct depressive types. A simple examination of the histogram of the scores for bimodality is a poor test, for the emergence of multimodality depends on the proportions, means, and standard deviations of the mixed subgroups or subpopulations (Everitt & Hand, 1981). A mixture analysis provides a way of estimating the mean and standard deviation of, and the proportion of total scores contributed by, each hypothesised subpopulation. We used the method described by Hasselblad (1966), and to test for the number of component subpopulations we used the likelihood ratio test examined by Everitt (1981*a*). A type I error rate of 0.01 was imposed, both to allow for multiple decisions, and to be conservative in our rejection of null hypotheses.

In developing refined sign and symptom indices of 'melancholia', we derived cut-off values for each index by examining the receiver operating characteristic (ROC) curve (Murphy *et al*, 1987) for a range of cut-off values using criterion clinical diagnoses of 'psychotic/endogenous' v. 'neurotic/reactive' depressive disorders. ROC curves plot false against true positives, with the points on the curve representing first increasing and then decreasing ability of the diagnostic test to discriminate.

Results

Refining signs

In the PCA, the first unrotated component in sample A accounted for 36.9% of the variance (with 19 items having loadings greater than 0.6) and in sample B it accounted for 40.6% of the variance (with 20 items having loadings greater than 0.6). Table I lists the loadings on this first component which in both samples strongly defined a retardation dimension. Oblique rotation of varying numbers of components consistently retained this dimension. The next four factors accounted for 9–4% (total 26.3%) of the variance in sample A, and for 10–3% (total 24.5%) of the variance in sample B. The variance data support the emphasis on the first component alone.

Table I also shows the results from the LCA. To dichotomise the four-point mental state items we decided on using 1 (feature absent) v. 2–4 (feature present to some degree). Ideally, the final items will have high true-positive rates, which are the coefficients for 'melancholia', and low false-positive rates, which are the coefficients under 'not melancholia'.

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

TABLE I

Factor loadings (first factor, pre-rotation) and probability estimates, using latent-class analysis of any individual occupying the latent melancholia and non-melancholia classes scoring positively on mental state signs, for sample A and sample B

Mental state sign	*Sample A (n = 262)*			*Sample B (n = 202)*		
	Loading on first factor in PCA	*Latent diagnostic group Melancholia*	*Non-melancholia*	*Loading on first factor in PCA*	*Latent diagnostic group Melancholia*	*Non-melancholia*
Uncooperative attitude	0.60			0.46		
Depressed affect initially	0.67			0.69		
Unresponsive to interviewer	0.69	0.83	0.22	0.77	0.86	0.16
Dull/inattentive	0.59	0.65	0.09	0.73	0.59	0.06
Depressed affect maintained	0.67			0.70		
Persistently worried	0.58			0.57		
Persistently miserable	0.66			0.70		
Fixed, immobile face	0.72	0.97	0.46	0.78		
Querulous and irritable	0.34			0.30		
Facial agitation	0.45			0.39		
Self-preoccupied	0.72	0.95	0.34	0.67	0.90	0.23
Inability to be cheered by interviewer	0.67			0.67		
Slumped posture	0.68	0.73	0.11	0.70	0.76	0.15
Immobility	0.64	0.72	0.07	0.74	0.78	0.06
Slowed movements	0.71	0.74	0.09	0.75	0.73	0.07
Slowed speech	0.65	0.68	0.06	0.70	0.72	0.06
Inaudible	0.51			0.67	0.68	0.04
Mute or reduced speech	0.71	0.79	0.09	0.71	0.71	0.05
Reduced spontaneous talk	0.36			0.79	0.86	0.11
Speech favouring self	0.52			0.65		
Reduced themes	0.36			0.71	0.81	0.22
Poverty of associations	0.72	0.93	0.42	0.76	0.94	0.25
Retardation of movement	0.43			0.66	0.74	0.12
Agitated movement	0.29			0.41		
Impaired concentration	0.17			0.59		
Impaired insight	0.71	0.87	0.17	0.50		
Excessive guilt	0.62			0.32		
Nihilistic	0.75	0.94	0.31	0.57		
Observable anxiety	0.77	0.93	0.28	0.33		
Endogenous quality	0.67	0.73	0.06	0.63		
Base rate (% in diagnostic group)		57%	43%		33%	67%

Table I shows that the initial LCA estimated that 57% of sample A belonged to the putative melancholia class, which agrees almost perfectly with the estimate by the clinician raters that 59% of this sample had 'psychotic' or 'endogenous' depression. In sample B, the estimates were less in agreement. The initial LCA gave an estimate of 33%, while the consultant psychiatrists had clinically diagnosed 50% as having psychotic or endogenous depression (including in the latter 33 patients with a mixed diagnosis of 'endogenous/neurotic' depression).

Finalising core signs and deriving core scores

We finalised our set of 'core' signs by including nine items suggested by both multivariate techniques and in both samples, and six items suggested strongly in one sample and/or by one multivariate technique.

Table II examines the percentage of subjects in each sample rated as showing some abnormality (i.e. scoring 2–4) on each mental state item, and whether those diagnosed clinically as having psychotic/endogenous depression were more likely to return a higher score on the four-point ordinal mental state scale, which we quantify with odds-ratio estimates. In broad terms, our final 'core' items were far more likely than non-core items to distinguish the psychotic/endogenous depression from remaining depressive disorders.

We then computed a 'core' score for each subject by summing raw (1–4) scores on each of those 15 items. Mean core scores were 27.8 (s.d. 8.5) in sample A (males 27.6, females 27.9), and 24.2 (s.d. 7.8) in sample B (males 23.3, females 24.6).

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

TABLE II

Percentage of samples rated as having some abnormality on mental state items, and comparison of scores returned by clinically differentiated subgroups in each sample, for core and non-core items

Mental state sign	*Sample A*			*Sample B*		
	Psychotic/ endogenous	*Neurotic/ reactive*	*Odds ratio*	*Psychotic/ endogenous*	*Neurotic/ reactive*	*Odds ratio*
Core items						
Unresponsive to interviewer	69	47	3.14***	54	30	2.72***
Dull/inattentive	51	22	3.74***	36	10	5.04***
Fixed, immobile face	89	61	4.92***	77	54	2.95***
Self-preoccupied	84	54	4.61***	55	41	1.75*
Inability to be cheered by interviewer	97	96	1.53	86	75	2.04
Slumped posture	56	25	3.81***	44	26	2.23**
Immobility	54	18	5.48***	38	18	2.78**
Slowed movements	58	19	5.88***	38	15	3.61***
Slowed speech	51	19	4.49***	35	16	2.82**
Mute or reduced speech	60	24	4.87***	35	16	2.82**
Poverty of associations	93	43	18.49***	61	42	2.06**
Impaired insight	74	33	5.97***	30	16	2.25*
Nihilistic	80	53	3.68***	34	19	2.19*
Observable anxiety	82	47	5.18***	78	74	1.25
Endogenous quality	57	16	7.03***	75	31	6.87***
Non-core items						
Uncooperative attitude	41	23	2.40**	24	19	1.35
Depressed affect initially	97	96	1.21	87	85	1.18
Depressed affect maintained	97	97	0.89	81	83	0.87
Persistently worried	86	80	1.53	95	92	1.65
Persistently miserable	91	87	1.45	78	79	0.94
Querulous and irritable	38	35	1.14	39	33	1.30
Facial agitation	58	50	1.39	61	51	1.56
Inaudible	38	22	2.24	30	15	2.42*
Reduced spontaneous talk	55	45	1.51	51	25	3.10***
Speech favouring self	49	28	2.47**	67	59	1.41
Reduced themes	71	75	0.82	53	36	2.00*
Retardation of movement	87	81	1.61	47	18	4.01***
Agitated movement	97	94	1.86	54	34	2.31**
Impaired concentration	34	41	0.75	72	38	4.32***
Excessive guilt	78	65	1.94*	58	56	1.08

*$P<0.05$, **$P<0.01$, ***$P<0.001$.

Determining the number of subpopulations reflected in core scores

We undertook mixture analyses to pursue this issue. The first hypothesis, that there was only one, not two subpopulations underlying the core scores, was rejected strongly in both sample A ($\chi^2=35.9$, $P<0.001$) and sample B ($\chi^2=64.5$, $P<0.001$). The second hypothesis that there were two, not three subpopulations was also rejected in both samples ($\chi^2=18.0$ and 18.3, $P<0.001$). The subsequent hypothesis, that there were no more than three subpopulations in each sample, was accepted ($\chi^2=7.2$ and 8.8) (Fig. 1). While there were differing proportions of subpopulations in each sample, their means and standard deviations were very similar, the means being 16.6 (sample A) and 17.7 (sample B) for the first subgroup, 23.6 and 23.9 for the second, and 34.4 and 36.8 for the third. Comparisons against clinical and other diagnostic systems suggested that the bulk of the clinically diagnosed 'neurotic/reactive' depressives belonged to the first subgroup, while most of the clinically diagnosed 'psychotic/endogenous' depressives belonged to the second and third groups, with the clinically 'agitated' patients more likely to be allocated to the second, and the clinically 'retarded' patients more likely to be allocated to the third.

Delineating melancholia by use of core scores

We judged then that 'melancholic depressives' were most likely to be in the second and third groups and sought to determine cut-off core scores for discriminating those with 'melancholia' by examining ROC curves, as described earlier. In sample A, the most discriminating cut-off core score was 24/25, while in sample B it was 20/21.

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

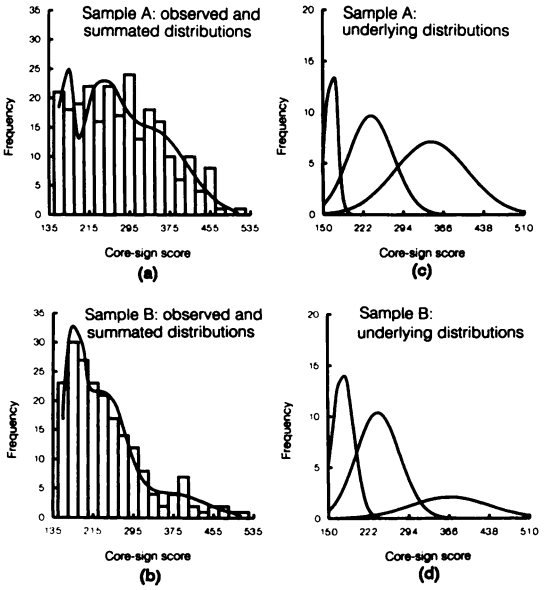


FIG. 1. a,b show histograms of the observed core-sign scores over which has been plotted the distribution obtained by summing the underlying distributions suggested by the mixture analyses. The latter are displayed in c,d. Thus, while visual inspection suggests unimodal distributions, the mixture analyses show how unimodality can arise from three relatively distinct distributions. All curves have been smoothed: in a,b, adjacent points on the x axis have been merged, and the range of the x axis artificially extended to enable a more accurate plot.

In sample A, 124 (83.8%) of the 148 clinically diagnosed psychotic/endogenous depressives scored above the cut-off point, in comparison to 79 (78.2%) of 101 in sample B. Of the remaining subjects (i.e. those with neurotic/reactive depression and personality disorder with depression), 69 (74.2%) of 93 from sample A (data missing for 21 subjects) and 58 (57.4%) of 101 from sample B scored below the cut-off point, giving a positive predictive power (Baldessarini *et al*, 1983) of 0.84 and 0.65, and a negative predictive power of 0.74 and 0.73 in each sample respectively.

We then examined mean core scores for subjects in each sample diagnosed according to different diagnostic systems: DSM-III and clinical diagnoses in each sample (Table III) and, additionally in sample B, RDC diagnoses (not shown in Table III). The RDC data were very similar to related DSM-III categories, with mean scores being 35.6 for the RDC psychotic, 25.2 for the endogenous, 20.3 for major, 19.7 for minor, 17.4 for intermittent, and 17.0 for situational depression. In each sample, and for each diagnostic system, the clinically diagnosed psychotic/endogenous depressives scored significantly higher than the remaining depressives, including the RDC system ($t=6.85$, $P<0.001$). Additionally, in sample B those patients assigned as endogenous by the Newcastle index had significantly higher core scores than remaining subjects. Table III also shows a general, but not consistent, trend for psychotic depressives to return higher core scores than endogenous depressives, again significant using RDC criteria ($t=5.95$, $P<0.001$).

Relationships between core scores and other variables

Older subjects returned higher core scores in both sample A ($r=0.35$, $P<0.001$) and sample B ($r=0.28$, $P<0.001$), and we then sought to determine whether this link was likely to be a direct one or a reflection of psychotic/endogenous depressives tending to be older. When diagnosis was partialled out from both core score and age, the correlation coefficient between age and core scores was reduced from 0.35 to 0.15 ($P<0.05$) in sample A, and from 0.28 to 0.18 in sample B. When subsamples were examined, the correlation between age and core scores existed in the psychotic/endogenous subsample ($r=0.17$, $P<0.05$ in sample A, and 0.27, $P<0.01$ in sample B) but not for the neurotic/reactive subsample ($r=0.10$ and -0.01, NS). Thus, we conclude that the link between core scores and age does not reflect a *general* phenomenon, although older subjects are more likely to have psychotic/endogenous depressive disorders.

Core scores were independent of sex ($t=0.24$ and 1.12 in samples A and B), of duration of depressive episode ($r=+0.06$ and -0.09 in samples A and B), but related to older age of onset of initial depressive episode ($r=0.27$) (examined in sample B only). Core scores were associated with both Newcastle ($r=0.52$, $P<0.001$) and HRSD ($r=0.42$, $P<0.001$) scores (examined in sample B only). The association with HRSD scores raises the possibility that core scores might merely be assessing severity to some degree. We judge that unlikely, however, as associations linking core with Zung ($r=0.04$) or GHQ ($r=0.06$) scores (both essentially self-report 'severity' scales) were not demonstrated. While the HRSD is a measure of severity, items are weighted to the concept of endogenous depression (principally retardation, agitation, psychotic features, pathological guilt, and loss of insight), so that associations between HRSD and core scores are more likely to reflect their capacity to address typological issues.

Despite difficulties in establishing validating factors for endogenous depression or melancholia (Andreasen, 1980), we undertook some analyses to determine whether core scores related to background or precipitating events. In regard to five specified acute stressors preceding onset of the current depression, 'high-core' subjects in sample B reported fewer stressors ($t=2.37$, $P<0.05$) and were particularly unlikely to report the break-up of an intimate relationship (8% v. 22%, $\chi^2=8.2$, $P<0.01$). The trend was for high-core subjects to be less likely to report being in a dysfunctional relationship (20% v. 30%, $\chi^2=2.82$, NS) and less likely to feel 'entrapped' by chronic stressors (28% v. 38%, $\chi^2=2.42$, NS).

In sample A, exposure to a set of antecedent stressors was reported in a self-report questionnaire assessing separate stressors (categorised as none, acute, chronic, or acute and chronic). Of subjects denying any stressors, 79% belonged to the high-core compared with 21% to the low-core group. For chronic stressors, high-core subjects were less likely to report on-going problems with their spouse or partner ($\chi^2=5.30$, $P<0.025$) and less likely to report a poor

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

TABLE III

Extent to which core and symptom scores differentiate clinical and other diagnosed subgroups in two separate samples

Diagnostic system and depressive category	*Sample A*			*Sample B*					
	No. of patients rated	*Mean core score*	*s.d.*	*No. of patients rated*	*Mean core score*	*s.d.*	*No. of patients rated*	*Mean symptom score*	*s.d.*
Clinical									
Psychotic depression[1]	24	37.1	7.4	15	37.9	9.4	12	10.8	2.3
Endogenous depression[2]	124	30.6	7.4	53	26.6	7.3	43	9.5	2.5
Neurotic depression	46	22.5	5.5	51	21.9	5.2	47	7.0	2.9
Reactive depression	47	21.2	6.0	9	19.2	3.2	8	5.9	3.0
Endogenous/neurotic				33	23.1	7.1	31	9.0	3.1
Reactive/neurotic				31	20.3	4.5	28	7.0	3.0
Personality disorder				10	21.7	5.1	10	8.3	4.3
DSM-III									
Major with psychotic features[1]	27	35.6	7.6	16	36.7	10.4	13	10.0	2.8
Major with melancholia[2]	44	32.6	6.9	61	26.9	7.2	54	10.2	1.8
Major + dysthymia	15	29.0	7.0		–			–	
Major	88	27.2	8.5	107	21.7	5.2	94	7.3	3.0
Dysthymia	35	24.1	6.0	15	18.2	3.2	15	5.3	3.2
Adjustment	36	20.7	4.5	2	16.5	2.1	2	7.5	9.2
Atypical				1	25.0		1	7.0	
Newcastle scale									
'Endogenous class'[1]				36	30.3	10.1	36	9.9	2.5
'Non-endogenous class'[2]				143	22.3	5.8	143	7.8	3.2

Note some subjects excluded from putative subgroups because of missing data.

Key contrast A: psychotic/endogenous subgroups[1,2] compared with remaining subgroups:

Signs – sample A DSM-III ($t=8.03$, $P<0.001$), clinical ($t=11.59$, $P<0.001$).

Signs – sample B DSM-III ($t=5.30$, $P<0.001$), clinical ($t=7.75$, $P<0.001$), Newcastle ($t=6.80$, $P<0.001$).

Symptoms – sample B DSM-III ($t=3.44$, $P<0.001$), clinical ($t=5.11$, $P<0.001$), Newcastle ($t=3.58$, $P<0.001$).

Key contrast B: psychotic subgroup[1] compared with endogenous subgroup[2]:

Signs – sample A DSM-III ($t=1.71$, NS), clinical ($t=4.31$, $P<0.001$).

Signs – sample B DSM-III ($t=5.51$, $P<0.001$), clinical ($t=6.03$, $P<0.001$).

Symptoms – sample B DSM-III ($t=0.26$, NS), clinical ($t=1.30$, NS).

work situation ($\chi^2=5.26$, $P<0.05$). Of subjects with a family history of depression, 74% had been assigned to the high-core group ($\chi^2=4.25$, $P<0.05$). Of the 21 patients who reported a previous manic attack, 17 (81%) had been allocated to the high-core group, a reassuring finding although, because of a low base rate, not formally significant ($\chi^2=3.77$, $P=0.05$).

As part of their clinical assessment, 29 patients in sample B had had a dexamethasone suppression test (DST). Non-suppression (using a cut-off of 138 nmol/l) occurred in 45% of high-core and none of the low-core subjects.

Refining symptoms

Next we sought to develop a refined 'endogeneity' symptom scale. Only sample B was studied as, in this group, a more comprehensive set of symptoms was elicited at interview by a clinician. LCA failed to produce a refined set of symptoms that were either clinically or historically compatible with the definition of endogenous depression. We then undertook a PCA. The first two components (Table IV) accounted for 14.4% and 8.5% of the variance. The first (i.e. as defined by the 15 items loading in excess of +0.30) is clearly compatible with published factorial definitions of 'endogeneity' or melancholia and, in part, again suggests a 'retardation' dimension, with anhedonia, non-reactivity, appetite and weight loss, together with slowed and indecisive thoughts dominating. The second factor was contributed to most strongly by appetite and weight gain, hypersomnia, and diurnal change in mood and energy. Odds-ratio calculations established that all 17 items loading more than +0.23 on the first factor were more likely to be affirmed by clinically diagnosed psychotic/endogenous depressives.

We then created a 'symptom' scale score by summing scores on the 14 symptoms loading highest on the first factor (Table IV). The next highest loading item (delusions, which were rarely affirmed) was excluded, as were 'agitation' and 'distinct mood' which, while having odds ratios suggesting their greater likelihood in the clinically diagnosed psychotic/endogenous subjects, had low factor loadings.

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

TABLE IV

Variables in principal-components analyses of symptoms, their pre-rotation loadings on the first two factors and their capacity to differentiate clinical diagnoses, in sample B

Variable	*Coding*	*Loadings*		*Percentage affirming symptom*		*Odds ratio*
		Factor 1	*Factor 2*	*Psychotic/ endogenous*	*Neurotic/ reactive*	
Consummatory anhedonia	B	0.65[1]	0.08	82%	71%	1.86
Anticipatory anhedonia	B	0.63[1]	0.35	79%	65%	2.02*
Appetite loss	marked	0.58[1]	−0.33	51%	21%	3.89***
Non-reactive in social situations	B	0.54[1]	0.41	74%	67%	1.40
Weight loss	marked	0.52[1]	−0.32	39%	26%	1.82
Non-reactivity (to likely major pleasant event)	B	0.47[1]	0.27	75%	67%	1.52
Indecisiveness	A	0.47[1]	0.10	83%	53%	4.51***
Preoccupied with unpleasant thoughts	A	0.44[1]	−0.05	58%	52%	1.27
Slowed thoughts	A	0.44[1]	0.25	87%	69%	3.01**
Terminal insomnia	A	0.43[1]	−0.26	51%	31%	2.40**
Non-variable mood	A	0.42[1]	0.10	57%	43%	1.82*
Loss of interest and pleasure	C	0.39[1]	0.23	81%	70%	1.82
Constipation	A	0.33[1]	0.07	47%	29%	2.16**
Retardation	A	0.33[1]	0.30	71%	45%	2.95***
Delusions	A	0.32[1]	−0.21	16%	2%	9.32**
Distinct mood		0.25	0.15	93%	72%	5.18***
Agitation		0.23	−0.26	47%	43%	1.19
Middle insomnia		0.15	−0.23	35%	38%	0.88
Suicidal thoughts	A	0.12	−0.01	54%	66%	0.58
Initial insomnia		0.12	−0.09	35%	32%	1.14
No energy in morning		−0.01	0.32	44%	27%	2.12*
Anger with others		−0.04	−0.01	18%	44%	0.28***
Mood worse in morning		−0.05	0.29	56%	41%	1.90*
Hypersomnia		−0.19	0.55	19%	26%	0.67
Weight gain		−0.20	0.59	16%	19%	0.81
Increased appetite		−0.32	0.54	18%	28%	0.57

A = yes/no or equivalent loading; B = less than 50%; C = more than 50%.
*$P<0.05$, **$P<0.01$, ***$P<0.001$.
1. Items selected to contribute to the 'symptom' scale.

Examining for links between core and symptom scores

The mean score for the 179 subjects (from sample B) with no missing data was 8.2 (s.d. 3.2), being 8.1 for males and 8.3 for females ($t=0.26$, NS). The correlation between 'core' sign and 'symptom' scores was significant ($r=0.39$, $P<0.001$). The presence of each of the 14 symptoms was associated with a higher score than its absence, although not all differences were significant. Discriminating symptoms, in terms of significantly higher core scores were: preoccupied with unpleasant thoughts ($t=4.76$, $P<0.001$); indecisiveness ($t=3.01$), non-variable mood ($t=2.96$), constipation ($t=2.88$), and retardation ($t=2.88$), all $P<0.01$; marked appetite loss ($t=2.49$), consummatory anhedonia ($t=2.31$), anticipatory anhedonia ($t=2.31$), non-reactivity in social situations ($t=2.16$), and slowed thoughts ($t=2.16$), all $P<0.05$, with significant associations suggesting that these symptoms again defined a retardation dimension.

Further assessment of symptom scores

In addition to being independent of sex, scores on the 'symptom' measure were unrelated to age ($r=0.07$), or to duration of depression ($r=-0.03$), but did show weak associations with age at first episode ($r=0.17$, $P<0.05$), with severity as assessed by the Zung ($r=0.27$, $P<0.001$) and the GHQ ($r=0.22$, $P<0.01$) scores, and moderate associations with scores on the HRSD ($r=0.35$, $P<0.001$) and Newcastle scale ($r=0.40$, $P<0.001$).

When derived 'symptom' scores were examined against separate diagnostic systems, differentiation was less marked than for core scores. Psychotic/endogenous depressives, as defined clinically and by DSM-III, returned higher symptom scores than remaining depressives, but psychotic and endogenous depressives returned similar symptom scores to each other (Table III). The same interpretations held for RDC-defined groups ($t=5.25$, $P<0.001$; $t=0.87$, NS respectively).

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

As for 'core' scores, we examined ROC curves for symptom scores against clinical diagnosis and determined a cut-off score of 7/8 in sample B. 'High-symptom' subjects so assigned were more likely to be in the 'high-core' subgroup, with an overall matching of 67.6%. By contrast, 'high-symptom' allocation was less strongly associated with an 'endogenous depression' diagnosis on the Newcastle scale, the overall matching being 54.7%. Examination of subgroups against risk variables established that 'low-symptom' subjects reported more of the five specified acute stressors ($t=3.62$, $P<0.001$), and were particularly likely to report a break-up in an intimate relationship (21% v. 9%, $\chi^2=4.91$, $P<0.05$), while in relation to chronic stressors, they were more likely to report a dysfunctional intimate relationship (37% v. 19%, $\chi^2=7.16$, $P<0.01$).

Core and symptom scores as comparative predictors of clinical diagnosis

We undertook a discriminant-function analysis to examine the comparative prediction of clinical diagnosis (psychotic/endogenous v. neurotic/reactive) by total symptom and total core scores, and established canonical coefficients of 0.66 (symptoms) and 0.59 (core), suggesting that symptoms and signs had a similar ability to predict clinical diagnosis, and with the overall classification rate being 68%.

Aware that our capacity to evaluate predictability was confounded by doubtful and mixed diagnoses, we undertook subanalyses of more clinically distinct groups. Specifically, we compared 26 subjects who were rated as 'definite' RDC psychotic or endogenous depression *and* as having a DSM-III major depression with melancholia or delusions *and* as 'endogenous' on the Newcastle scale, in comparison with 47 subjects who had a 'definite' RDC major depression *and* a DSM-III major depression or adjustment disorder *and* were rated as 'non-endogenous' on the Newcastle scale. A discriminant-function analysis, entering each of the 15 core signs and each of the 14 symptoms, had an overall correct classification of 91.8% (and a distinct bimodal distribution). More importantly, signs entered alone reduced Wilks' lambda to 0.39 ($F=5.87$, equivalent to $R^2=0.61$) whereas symptoms entered alone reduced lambda only to 0.55 ($F=3.40$, $R^2=0.45$), suggesting that, overall, signs were a stronger discriminator than symptoms. After entering signs first, the entry of symptoms reduced Wilks' lambda by a further 0.18. By contrast, entry of signs after entering symptoms first reduced lambda by a further 0.33. These analyses suggest that, in clinically distinct subgroups (with diagnostically difficult or mixed-diagnosis patients excluded), our derived signs are stronger predictors than the derived symptoms and that, after shared variance is accounted for, core signs make an independent contribution distinctly greater than that made by symptoms.

Reliability

Using all subjects, the internal consistency (α) of the 15 high-core items was examined, and coefficients of 0.93 in sample A and 0.91 in sample B were established, demonstrating an extremely homogenous dimension underlying core items. The internal consistency of the 14 symptoms refined in sample B was 0.76, suggesting less homogeneity.

More importantly, we formally tested the inter-rater reliability of the derived core items by having 10 patients rated by MDU consultants working in pairs or groups. For total core scores, the inter-rater correlations ranged from 0.76 to 0.98 for the separate pairings. These preliminary data suggested that without training or operational definitions, clinicians achieved very high consensus in total core scores.

Discussion

In our attempt to differentiate those with 'melancholia' from a residual group, we present data advocating the utility of assessing signs, having earlier noted some limitations of symptom-favoured approaches. Although the assessment of signs has some limitations, being particularly dependent on ratings made when the patient is significantly depressed, and perhaps dependent on time of day, and on effects of concurrent medication, electroconvulsive therapy (ECT), and physical illness, we have shown that high inter-rater reliability for total core scores exists despite a lack of firm operational criteria at this stage of the research.

In the absence of 'gold-standard' validating criteria, delineation of depressive types may proceed in two broad ways. A 'top-down' strategy compares clinical features of those assigned to certain classes. There is an immediate problem – the classification may not be valid. A 'bottom-up' strategy tries to isolate allied categorical or dimensional characteristics and most attempts to derive typologies have thus proceeded using multivariate techniques such as factor and cluster analysis. The problem here is validating the set of features derived. Andreasen (1980) has advocated the demonstration of distinct aetiological factors, distinct clinical course, or a distinct response to specific treatments (predictive validity), while Carroll (1984) has proposed the use of laboratory markers. In fact, firm and consistent validation has yet to be achieved for any typological system.

A clear retardation dimension or construct was evident in both the derived signs and symptoms. Those with a classic agitated depression (but clinically diagnosed as endogenous or psychotic) also tended to return high 'core' scores, but not as high as classic retarded patients, presumably because of the bias to retardation in the item descriptions. As 'agitation' and 'retardation' tend to load on the same factor in most factor-analytic studies (see Mendels & Cochrane, 1968; Nelson & Charney, 1981), and as Widlöcher

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

(1983) noted that "even grossly agitated patients have been shown to have retardation (for example, prolonged speech pause ›time)", our finding is consistent with previous studies. Previous reviews of these factor-analytic studies have demonstrated the central importance of retardation. Nelson & Charney (1981), for instance, stated that "Psychomotor change is the symptom most strongly associated with autonomous depression. Retardation was especially consistent in the factor analytic and drug response studies", while retardation again "received strong support" in their review of cluster-analytic studies.

In his definitive review of melancholia, Lewis (1934) noted that retardation has long been recognised by clinicians as an important feature of depressive states, and he reviewed its component constructs in his sample, as well as reviewing European concepts of retardation, including Wilmanns' view that intrapsychic inhibition is "characteristic of melancholia with its anxious and attentive self-observation" and that "no less an authority than Wernicke has attributed to the retardation of thought a predominant significance in melancholia." Widlöcher (1983) also noted historical descriptions of a general slowing down of motor activity (arms often fixed, trunk immobile, expression unchanging with an immobile and blank face) and of psychic activity (slow responses, lack of spontaneity, impoverishment of new themes, brief and laconic responses), often with poor concentration, an apparent memory disorder, ruminative thinking and a sense of fatigue and distress, while speech is said to be affected in both motor and cognitive aspects (weak and monotonous voice, slow verbal flow, brief responses).

Widlöcher suggested that retardation is a core behavioural pattern and a primary disturbance in affective disorders. Such a view is not surprising if retardation and severity are equated, as is frequently proposed (Ní Bhrolcháin, 1979). As noted earlier, most reviews have included severity as a variable favouring a diagnosis of endogenous depression, while classificatory systems such as DSM–III weight a diagnosis of major depression very much on the basis of severity. What is particularly striking about our derived core items is that they appear to be independent of severity (in comparison with symptoms, which were weakly related to severity) and suggest that melancholia can be defined phenomenologically without reference to severity.

We need to consider the possibility of a halo effect, whereby a rater might make a clinical diagnosis and *then* complete ratings of signs so as to support such a diagnosis. One item in particular ('endogenous quality') is a likely candidate for such an explanation. A halo effect is unlikely to have determined the general study findings as significant odds ratios were not always associated with features said to be characteristic of psychotic/endogenous depression. Nevertheless, we shall examine for any such effect in subsequent studies, particularly by use of corroborative witness reports. More importantly, in addressing the potential for clinicians to make a diagnosis using the same mental state signs as refined for the core, we note that in sample B, DSM–III and RDC diagnoses were made in strict accordance with operational criteria and that such symptom-focused diagnoses were derived independently of core signs and yet generated differential core-sign scores across putatively differing diagnostic categories.

Bimodality of scores has been held (Kendell, 1975; Garside & Roth, 1978; Everitt, 1981*b*) to be a necessary, but not sufficient, indicator of disease categories, a view criticised by Grayson (1987). Nevertheless, we formally tested whether core scores had a unimodal distribution, or were suggestive of more than one underlying subpopulation. Mixture analyses suggested three subpopulations with similar means and standard deviations in each sample. As the mean core score in the first subpopulation in each sample only slightly exceeded the possible minimum, it clearly comprised those scoring at the base of the core scale, be it as a consequence of trivial depression (unlikely when we excluded low scores on the Zung scale) or the complete absence of core signs. The second and third subpopulations appeared to reflect 'agitated' and 'retarded' versions of melancholia. In refining our mental sign measure we shall seek greater potential equivalence of scores for 'agitated' and 'retarded' depressives by including more agitation items.

If our core signs have the capacity to distinguish melancholia from residual depressive disorders we would expect some differentiation against background variables. While a number have been suggested (Zimmerman *et al*, 1986), none has been clearly established in empirical studies (Andreasen, 1980), and thus they are no more than weak validators. Additionally, our analyses (like most studies of this nature) were based on data not assessed in great detail (e.g. life events) and, at times, on subsamples (e.g. DST). Findings, nevertheless, were again encouraging. Thus, across samples, subjects returning a high core score were more likely to report no or fewer antecedent life stressors, to report delusions and hallucinations in the present episode, to have a family history of depression, and to be older at onset of their initial depression, all variables that have been indicated as more often present in melancholia or endogenous depression.

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press

Additionally, core scores were related to clinical, DSM-III, and RDC diagnoses, and to scores on the Newcastle scale – an assumed measure of 'endogeneity'.

The present system and a projected revision (adding some agitation items and using precise descriptors and refined operational criteria) should be of particular use to researchers who will be able (as in the present study) to rate depressives on symptoms, signs, and on a composite index, to provide a probability estimate of melancholia. For the clinician, we seek to improve current systems and to narrow the boundary of mixed and doubtful diagnoses by deriving cut-off scores allowing more confident estimates, as well as to quantify the predictive utility of the system in estimating response to broad therapeutic approaches. In particular, we shall assess whether response to antidepressant medication and ECT is best predicted by signs, symptoms, or the composite index.

Acknowledgements

These studies were supported by two grants from the National Health and Medical Research Council (NH & MRC, Australia) and by a Coast Centenary Research grant from Prince Henry Hospital. We were assisted considerably by Jenny Appleton, Heather Brotchie, Catherine Hickie, Sandra Evans, Christine Taylor, Rhondda Hartley, Wayne Hall, Elizabeth Murray, Margaret Cooney, and Noel Uhe, as well as the many patients and Sydney psychiatrists who generated data.

*Gordon Parker, MD, PhD, FRANZCP, *Professor of Psychiatry, University of New South Wales*; Dusan Hadzi-Pavlovic, BSc, MPsychol, *Scientific Officer, Division of Psychiatry, Prince Henry Hospital*; Philip Boyce, FRANZCP, *Senior Lecturer, School of Psychiatry, University of New South Wales*; Kay Wilhelm, FRANZCP, *Staff Specialist, Division of Psychiatry, Prince Henry Hospital*; Henry Brodaty, MD, FRACP, FRANZCP, *Senior Staff Specialist, Division of Psychiatry, Prince Henry Hospital*; Philip Mitchell, MRCPsych, FRANZCP, *Senior Lecturer, School of Psychiatry, University of New South Wales*; Ian Hickie, FRANZCP, *Staff Specialist, Division of Psychiatry, Prince Henry Hospital*; Kerrie Eyers, MA, MAPS, *Research Psychologist, Division of Psychiatry, Prince Henry Hospital*

*Correspondence: *Mood Disorders Unit, Prince Henry Hospital, Little Bay 2036, Sydney, Australia*

References will be found at the end of the second paper, pp. 69–71.

British Journal of Psychiatry (1990), **157**, 65–71

Prediction of Response to Electroconvulsive Therapy
Preliminary Validation of a Sign-Based Typology of Depression

IAN HICKIE, BRIAN PARSONAGE and GORDON PARKER

In an open study of 35 depressed patients, psychomotor disturbance, assessed by the CORE rating system, significantly predicted response to ECT in patients with a 'retarded' but not those with an 'agitated' type of illness. In multivariate analyses, psychomotor disturbance was the sole predictor of response. This preliminary study suggests that the CORE system, which has been shown to differentiate endogenous and non-endogenous depressive disorders, has predictive validity in relation to ECT response.

Considerable debate continues about the subclassification of depressive disorders, with phenomenological characteristics, symptom clusters, mathematical models, biological markers, and specific treatment response variably proposed as the most useful strategies. While any classificatory system may have intrinsic merit, clinical utility is a more important requirement, in its capacity both to specify treatment modality and to predict outcome.

In addressing the subclassification of depressive

https://doi.org/10.1192/bjp.157.1.55 Published online by Cambridge University Press