

# Measuring melancholia: the utility of a prototypic symptom approach

G. Parker<sup>1,2\*</sup>, K. Fletcher<sup>1,2</sup>, M. Hyett<sup>2</sup>, D. Hadzi-Pavlovic<sup>1,2</sup>, M. Barrett<sup>2</sup> and H. Synnott<sup>2</sup>

<sup>1</sup> School of Psychiatry, University of New South Wales, Sydney, Australia

<sup>2</sup> Black Dog Institute, Sydney, Australia

**Background.** Melancholia has long resisted classification, with many of its suggested markers lacking specificity. The imprecision of depressive symptoms, in addition to self-report biases, has limited the capacity of existing measures to delineate melancholic depression as a distinct subtype. Our aim was to develop a self-report measure differentiating melancholic and non-melancholic depression, weighting differentiation by prototypic symptoms and determining its comparative classification success with a severity-based strategy.

**Method.** Consecutively recruited depressed out-patients ( $n=228$ ) rated 32 symptoms by prototypic or ‘characteristic’ relevance (using the Q-sort strategy) and severity [using the Severity-based Depression Rating System (SDRS) strategy]. Clinician diagnosis of melancholic/non-melancholic depression was the criterion measure, but two other formal measures of melancholia (Newcastle and DSM-IV criteria) were also tested.

**Results.** The prevalence of ‘melancholia’ ranged from 20.9% to 54.2% across the subtyping measures. The Q-sort measure had the highest overall correct classification rate in differentiating melancholic and non-melancholic depression (81.6%), with such decisions supported by validation analyses.

**Conclusions.** In differentiating a melancholic subtype or syndrome, prototypic symptoms should be considered as a potential alternative to severity-based ratings.

Received 15 February 2008; Revised 13 May 2008; Accepted 29 July 2008; First published online 16 September 2008

**Key words:** Depression, melancholia, prototypic, Q-sort strategy, severity.

## Introduction

... melancholia may be a naturally occurring phenotype, qualitatively distinguishable from non-melancholic depression. (Leventhal & Rehm, 2005)

One of the longest-standing controversies in psychiatry involves the classification of depression. The possibility exists that, within the overall set of depressive conditions, there is a depressive subtype (variably termed endogenous, endogenomorphic, autonomous, vital and melancholia over time) that has resisted definition for a number of reasons. First, many suggested markers (Parker & Hadzi-Pavlovic, 1996; Taylor & Fink, 2006) may measure depression *per se* rather than being specific to melancholia, with Nelson & Charney (1981) identifying few features with distinct specificity. Second, apart from the possible exception of psychotic features, depressive symptoms are imprecise and do not approach the optimum criterion of being ‘necessary and sufficient’ to allow

categorical definition. Most symptoms vary in severity and are subject to a range of self-reporting biases (e.g. denial, minimizing or magnification).

Such problems necessarily limit the capacity of any measure (especially a severity-weighted one) to delineate and measure melancholia, and the standard approach is to undertake a ‘semi-structured interview’ (Leventhal & Rehm, 2005), with DSM-IV criteria being ‘the most frequently used’ reference point. Few other measures have been developed. The Bech–Rafaelsen Melancholia Scale (MES; Bech & Rafaelsen, 1980) is one of few symptom measures, developed as an extension of the Hamilton Depression Scale. However, Smolka & Stieglitz (1999), as well as Bech (2002), state that the MES is more a measure of depression severity than of melancholia. Other symptom-based measures of melancholia operationalize DSM criteria, such as the Inventory to Diagnose Depression (IDD; Zimmerman & Coryell, 1987) and the self-report version of the Inventory of Depressive Symptomatology (IDS-SR; Rush *et al.* 2003). A rare example of a self-report measure designed to discriminate melancholic depression is the Levine–Pilowsky Depression Questionnaire (LPDQ; Pilowsky *et al.* 1969), which was

\* Address for correspondence: Professor G. Parker, Black Dog Institute, Hospital Road, Randwick, NSW 2031, Australia.  
(Email: g.parker@unsw.edu.au)

developed within a sample including 38 patients diagnosed with 'endogenous depression' and 38 patients with 'neurotic depression' but which did not return particularly impressive overall correct classification rates in that sample.

In terms of composite illness measures, DSM-III-R criteria included three non-symptom criteria (i.e. no significant antecedent personality disturbance; previous depressive episodes followed by remission; and previous good response to physical treatments). The Newcastle Scale (Carney *et al.* 1965) comprises symptoms, an illness course variable and also an assessment of personality 'adequacy' and precipitating events. However, measurement of any syndrome or entity is preferably limited to symptoms rather than incorporating illness course and treatment response variables, which are better viewed as validators.

In our earlier studies (Parker & Hadzi-Pavlovic, 1996) we considered the comparative capacity of sets of 'endogeneity symptoms' and observable CORE signs of psychomotor disturbance to define and differentiate melancholic depression. The 18-item CORE measure was superior to the symptom measure in differentiating 'melancholia', and superior to endogeneity symptom scores across a range of validation strategies involving psychosocial and biological variables, predicting response to both antidepressant drug therapy and electro-convulsive therapy (ECT). The greater discrimination of 'signs' encouraged us to argue that melancholia required observational rating of psychomotor disturbance.

However, two limitations have emerged over time. First, observable psychomotor disturbance is less distinctive in younger patients with seemingly 'true melancholia'. Second, valid rating of psychomotor disturbance requires observing patients at or near nadir of their depressive episode, a limitation in clinical practice when many patients do not present for diagnosis at their worst.

We therefore sought to develop a symptom-based measure in an attempt to overcome such limitations, and judged that there were four key issues to address. First, we aimed to improve on previous candidate sets of potentially specific melancholic symptoms. As 'psychomotor disturbance' is the most consistently identified marker of melancholia (Rush & Weissenburger, 1994), we introduced seven items capturing both motoric and concentration components of psychomotor disturbance. Second, we included those 'endogeneity' symptoms most consistently identified in previous reviews. Third, although the heterogeneous nature of non-melancholic depression ensures that it has no definable symptoms, we tested the utility of certain symptoms (e.g. anger, irritability) that are more likely to be reported by those with non-melancholic

disorders. Fourth, we needed to overcome limitations to most depression severity self-report measures, where ratings may reflect individual response biases (Demyttenaere & De Fruyt, 2003) rather than capture true 'severity', and with such biases more likely in non-endogenous and dysthymic patients (Rush *et al.* 1987). We therefore sought to rate symptoms in terms of their 'prototypic' status, adopting a Q-sort strategy. By also including a standard Likert-type scale rating severity of the same set of symptoms, we could quantify the comparative discrimination offered by each strategy.

Lacking any 'gold standard' reference for any melancholic *versus* non-melancholic depression decision, we used 'clinician diagnosis' as the reference measure, with two clinicians making independent ratings, and with the senior clinician essentially adopting a 'Longitudinal, Expert, All Data' (LEAD; Spitzer, 1983) strategy to decision making, effectively by assessing longitudinal and cross-sectional information from multiple sources. Finally, we report several validation analyses of the measure.

## Method

### General

Patients referred to our out-patient Depression Clinic for diagnostic and management advice completed a computerized Mood Assessment Program (MAP), which collects (by self-report and clinician-rated) sociodemographics, illness details (including assessing features of bipolar disorder), personality and treatment details, and lifetime rates of anxiety disorders. MAP data included: (i) sociodemographic (age, sex, marital status) data; (ii) current depressive disorder characteristics [duration of episode, and severity as assessed by two state depression measures, the 10-item depression in the medically ill (DMI-10; Parker *et al.* 2002) and the Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR; Rush *et al.* 2003)]; (iii) depressive disorder history (age of onset of first depressive episode); (iv) family history of depression, bipolar disorder and alcoholism; (v) developmental difficulties with parents; (vi) stressful life events during the preceding 12 months; (vii) previous receipt of ECT and its self-reported level of effectiveness; (viii) overall functioning limitations (clinician-rated and patient-rated) or disability levels; and (ix) ongoing and pre-episode personality functioning limitations (clinician-rated).

MAP administration was followed by a semi-structured interview assessing DSM-IV criteria for major depressive disorder and melancholia. This interviewer (K.F.) was available for 187 (82%) of the

patients and blind to MAP diagnostic decisions and all referral information.

The intake psychiatrists (M.B. and H.S.), both of whom have had extensive clinical experience in assessing those with mood disorders, then undertook a detailed clinical assessment for 60–90 minutes, derived Newcastle Scale ratings (cut-off score  $\geq 6$  assigning melancholic depression) subject to interview time constraints, and subsequently presented the history to the senior psychiatrist (G.P.). Here, documentation from referring physicians and any other source was reviewed, followed by the senior psychiatrist interviewing the patient (and often a relative) to clarify diagnostic issues. All psychiatrists were blind to MAP diagnostic decisions, with the senior psychiatrist also blind to Newcastle Scale results. As noted, we therefore collected longitudinal and cross-sectional data from patients, from their referring physicians and often from multiple sources, seeking to adopt a LEAD strategy.

Although the clinic is a tertiary referral centre, the majority of patients were referred by general practitioners for diagnostic and management clarification, and only a minority had a referral from a psychiatrist for a treatment-resistant depression. Nevertheless, such referral nuances may have weighted the sample to a more diagnostically difficult group than if recruitment had occurred in a general practice or a general psychiatrist setting.

#### Study-specific nuances

Q-sorting is an 'ipsative' technique whereby the 'sorter' is required to examine each item in relation to every other item presented. By using a forced distribution, the sorter is restricted to limited grid positions in which to place symptoms, thus addressing the limitation of Likert-type scales, where all items may be inflated or minimized by a rater bias (Block, 1961). For a detailed discussion of Q-sort methodology, the reader is referred to the article by Watts & Stenner (2005).

Q-methodology ordinarily requires 40–80 items (Watts & Stenner, 2005), with some researchers adopting 100 or more items in some instances (e.g. Jones, 1985). Large numbers of items are burdensome (particularly for depressed individuals), whereas smaller item sets risk inadequate coverage of the area of interest. We sought to maximize the validity of the methodology in our sample by using a forced-normal distribution (Fig. 1) that minimizes the potential for subjects to rank symptoms at the extreme points, an aspect that we believe is particularly salient for the population of interest, who are noted to often 'rate up' symptom severity to emphasize the seriousness of



Fig. 1. Final Q-sort grid template, illustrating the template allowed for patients to allocate prescribed numbers of 'least characteristic' through 'neutral' to 'most characteristic' features.

their condition. The forced-normal distribution dictates the number of items that can be assigned to each ranking position. Bearing these issues in mind, and respecting the need for an approximately equal number of 'melancholic' and 'non-melancholic' symptoms to reduce potential bias, we wrote 32 descriptors (see Table 1) that we judged (from historical and our own clinical studies) as having some specificity to melancholic or to non-melancholic depression. The same 32-item set was presented for self-report completion using the computerized MAP, first using the Q-sort format and then the Severity-based Depression Rating System (SDRS) to generate Q-sort scores and SDRS scores.

Patients were instructed to focus on when they were at their 'worst' of their depression, to ignore medication side-effects and to effectively rank items from 'extremely characteristic' to 'extremely uncharacteristic'. They used a computer mouse to assign items and generate a final grid template comprising the two 'most characteristic' items (scoring +4), the next three 'most characteristic' (+3), the next four 'most characteristic' (+2), the next four 'most characteristic' (+1), and also a similar grid for the least characteristic items (respectively scoring -4, -3, -2 and -1) and with six 'neutral' items (scoring 0) also selected. The computer strategy allowed items to be progressively assembled, reviewed and altered by the patient before their final confirmation, as shown in Fig. 1.

Patients were subsequently presented with the same 32-item set (SDRS measure) and asked to rate whether they had experienced each symptom severely, moderately, mildly or not at all (scored 3, 2, 1 and 0 respectively), in relation to the same reference period for their episode.

Without any reference to the MAP, both the assessing and the senior psychiatrist (G.P.) were required to

**Table 1.** Item mean scores for clinically diagnosed ('extremely confident') melancholic and non-melancholic patients ( $n=74$ )

Item	N-			Item	N-		
	Mel	Mel	Mean score difference		Mel	Mel	Mean score difference
Feeling physically slowed	5.89	4.05	+1.84	Energy worse in morning	4.95	4.51	+0.44
Greater need to be close to people	4.79	3.20	+1.59	Loss of interest in enjoyable activities	6.58	6.33	+0.25
Depressed mood worse in morning	6.37	5.20	+1.17	Inability to be cheered up by friends	4.68	4.44	+0.25
Indecisiveness	7.00	5.95	+1.05	Brain feeling foggy, concentration difficult	6.00	5.98	+0.02
Loss of capacity to laugh	5.84	4.82	+1.02	Energy worse in evening	4.32	4.56	-0.25
Thinking slowed	5.74	4.85	+0.88	Early morning wakening	4.32	4.58	-0.27
'Paralysed' when doing basic things	5.74	4.91	+0.83	Appetite increase	3.47	3.96	-0.49
Appetite decrease	4.79	4.02	+0.77	Weight gain	3.37	3.93	-0.56
Unable to obtain pleasure	6.42	5.71	+0.71	Depressed mood worse in evening	3.68	4.31	-0.62
Inability to be cheered up by something nice	5.89	5.25	+0.64	Waking in middle of night	4.53	5.18	-0.66
Cheered up completely by friends	4.32	3.69	+0.62	Difficulty going to sleep	4.21	5.18	-0.97
Weight loss	4.00	3.44	+0.56	Need to stay away from people	5.58	6.84	-1.26
Agitated	5.26	4.76	+0.50	Deserve to be punished	3.16	4.60	-1.44
Difficulty doing basic things (e.g. get out of bed, shower)	5.79	5.29	+0.50	Guilt	4.26	5.82	-1.56
Unable to look forward to pleasurable activities	6.47	5.98	+0.49	Irritable	4.53	6.25	-1.73
				Suicidal	4.32	6.29	-1.98
				Angry	3.74	5.93	-2.19

Mel, Melancholic depression; N-Mel, non-melancholic depression.

+ indicates higher mean score in melancholic subset; - indicates higher mean score in non-melancholic subset.

make a judgment about whether the patient had either a melancholic or a non-melancholic depressive disorder, and rate the confidence of their diagnosis (5 = extremely confident, 4 = very confident, 3 = neutral, 2 = not very confident, 1 = not at all confident).

### Sample characteristics

The provisional sample comprised 234 patients clinically diagnosed with a unipolar depressive episode at the Institute's Depression Clinic during 2005–2007. Of those assessed, six were removed from the data analysis because of invalid self-report data as determined by the assessing clinician, generally reflecting distinctly impaired concentration or memory.

## Results

### Study samples

Analyses were undertaken on the whole sample ( $n=228$ ) and on a subset of 74 patients judged by the senior psychiatrist as allowing an 'extremely confident' subtyping decision of either melancholic ( $n=86$ ,  $n=19$  for total and subset samples respectively) or non-melancholic ( $n=142$ ,  $n=55$  for total and subset samples respectively) depression. Table 5 details

further sample characteristics including mean age and gender distribution.

Mean depression severity for the whole sample was calculated as 14.6 (S.D. = 5.4) for the QIDS-SR and 19.7 (S.D. = 7.6) for the DMI-10. Agreement between the independent allocations (melancholic *versus* non-melancholic) made by the assessing and the senior psychiatrist was high for both the whole sample ( $\kappa=0.78$ ) and the 'extremely confident' subset ( $\kappa=0.89$ ).

### Item testing

Q-sort scores returned by those with clinically diagnosed melancholic and non-melancholic depression were calculated in the 'extremely confident' subset, and items ranked (see Table 1) in terms of their diagnostic weighting and differentiation. Those with melancholia were more likely to prioritize items assessing psychomotor disturbance, mood and energy worse in the morning, anhedonia and mood non-reactivity, whereas those with non-melancholic depression were more likely to prioritize anger and irritability, guilt and feeling suicidal.

Our principal analyses were undertaken on the whole sample of 228 patients rather than being restricted to 'clear-cut' cases. Q-sort symptom scores for the 32 items were entered into a logistic regression,

**Table 2.** Area under the curve (AUC) comparison for SDRS and Q-sort measures

Sample		Test	AUC	S.E.	95% CI
Whole	1	Q-sort	0.855	0.027	0.802–0.909
	2	SDRS	0.856	0.025	0.807–0.906
‘Extremely confident’	3	Q-sort	0.910	0.034	0.844–0.976
	4	SDRS	0.895	0.039	0.818–0.972

  

Comparison	Difference	S.E.	95% CI	p value
1 v. 2	–0.001	0.025	–0.049 to 0.047	0.973
3 v. 4	0.015	0.039	–0.061 to 0.091	0.394

SDRS, Severity-based Depression Rating System; S.E., standard error; CI, confidence interval.

with the senior psychiatrist’s diagnosis (i.e. melancholic *versus* non-melancholic depression) as the independent grouping variable, to calculate an overall Q-sort test score. For comparative purposes, the analyses were repeated using the 32-item SDRS scores. The logistic regression model for the Q-sort measure correctly classified 81.6% of the total sample (72.1% of melancholic patients and 87.3% of non-melancholic patients) and 85.1% of the ‘extremely confident’ subset (73.7% melancholic and 89.1% non-melancholic) when compared to senior Clinician diagnosis. The SDRS yielded overall correct classification rates of 77.6% (65.1% melancholic and 85.2% non-melancholic) in the total sample and 77.0% (68.4% melancholic and 80.0% non-melancholic) in the ‘extremely confident’ sample.

Receiver operating characteristic (ROC) analyses were undertaken on Q-sort and SDRS test scores to evaluate their comparative diagnostic subtyping propensities as quantified by the area under the curve (AUC) of both measures. Table 2 lists AUC values for both measures and tests for pair-wise differences. The highest AUC was produced by the Q-sort measure within the diagnostically ‘extremely confident’ subset; however, no significant differences were demonstrated across samples or measurement strategies.

The optimal cut-off score for each measure was calculated as corresponding to the maximum of Kraemer’s (1992) QROC criterion  $\kappa$  (0.5, 0), weighting false positives and negatives equally. QROC- and ROC-derived cut-off values, sensitivity, specificity, positive and negative predictive values are presented in Table 3.

Examination of the performance of both measures within the total sample quantified the Q-sort measure as having marginally higher specificity values than the SDRS measure (regardless of whether QROC or ROC

cut-off scores were used). For the ‘extremely confident’ subset, the Q-sort strategy had higher sensitivity values than the SDRS strategy, whereas the SDRS strategy returned slightly higher specificity values. In terms of the positive predictive power (i.e. correctly allocating ‘true cases’ of melancholia), the Q-sort strategy was only superior to the SDRS within the whole sample set. Overall, the Q-sort strategy had superior negative predictive power (i.e. correctly allocating ‘true cases’ of non-melancholic depression) to the SDRS in both samples. As an ideal screening measure should have high specificity and negative predictive values, we suggest that the Q-sort strategy is marginally superior to the SDRS as a potentially useful screening tool based on the values derived in this sample.

#### *Assignment and agreement between the differing subtyping strategies*

The five differing systems assigned varying percentages to a melancholic category: 54.2% by DSM-IV, 37.7% by Clinician, 35.1% by Q-sort, 33.8% by SDRS and 20.9% by the Newcastle Scale. Table 4 examines the overall diagnostic agreement between each of the various measures, limited to the subset of 107 subjects where we had complete DSM-IV data.  $\kappa$  coefficients indicate that DSM-IV assignment was minimally in agreement with all other measures, that Newcastle assignment was slightly associated with Q-sort and SDRS and moderately so with Clinician diagnosis, and that Q-sort, SDRS and Clinician diagnosis were all moderately associated with each other.

#### *Validation analyses*

Several ascriptions to melancholia have been detailed (Parker & Hadzi-Pavlovic, 1996; Taylor & Fink, 2006),

**Table 3.** Receiver operating characteristic (ROC) analyses for SDRS and Q-sort measures

Sample	Test	Cut-off	Value	Se	Sp	PPV	NPV
Whole	Q-Sort	ROC	≥0.360	0.651	0.937	0.862	0.816
		QROC	≥0.360	0.651	0.937	0.862	0.816
	SDRS	ROC	≥0.198	0.628	0.887	0.771	0.797
		QROC	≥ -0.563	0.814	0.761	0.673	0.871
'Extremely confident'	Q-Sort	ROC	≥0.097	0.737	0.909	0.737	0.909
		QROC	≥ -0.381	0.842	0.873	0.696	0.941
	SDRS	ROC	≥1.170	0.579	0.982	0.917	0.871
		QROC	≥1.170	0.579	0.982	0.917	0.871

SDRS, Severity-based Depression Rating System; Se, selectivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value.

**Table 4.** Diagnostic agreement ( $\kappa$  values) between differing diagnostic strategies (whole sample)

	Q-sort <sup>a</sup> (n = 228)	SDRS <sup>a</sup> (n = 228)	Newcastle Scale (n = 187)	DSM-IV (n = 107)	Clinician (n = 228)
Q-sort		0.57***	0.40***	0.21*	0.53***
SDRS			0.32***	0.19*	0.44***
Newcastle Scale				0.16*	0.62***
DSM-IV					0.26**

SDRS, Severity-based Depression Rating System.

$\kappa$  values refer to overall diagnostic agreement between measures.

<sup>a</sup> Diagnostic decisions were derived by logistic regression analysis of the 32-item set.

\* Significant at the  $p < 0.05$  level (two-tailed). \*\* Significant at the  $p < 0.01$  level (two-tailed). \*\*\* Significant at the  $p < 0.001$  level (two-tailed).

including an older age, older age at first onset, greater severity, a more 'endogenous' onset (as against being reactive to life event stressors), exposure to fewer developmental stressors in childhood including dysfunctional parenting, being less likely to be associated with a personality disorder, a stronger family history of depression and a more specific response to physical treatments, particularly ECT.

Table 5 examines the degree to which assignment to melancholic depression by the five differing systems could be supported by several validating variables. Those assigned as melancholic by all five systems had a higher mean current age and, apart from DSM-IV diagnoses, a higher age of onset of depression, and tended to have longer episodes than non-melancholic subjects. Those assigned as having melancholic depression by all systems apart from DSM-IV were less likely to report difficulty with parents in their developmental years (significant for Q-sort, SDRS and Clinician) and fewer stressful life events in the

preceding 12 months. Q-sort, Newcastle and Clinician-assigned melancholic patients were significantly less likely to be rated with personality dysfunction. Receipt of ECT was more likely for those assigned as having melancholic depression by all systems other than by DSM-IV criteria, but no system identified any self-reported differential efficacy to ECT across those with melancholic or non-melancholic allocations.

**Melancholia: do study items support a separate type?**

According to Kendell (1989), a bimodal distribution is indicative of the existence of two discrete depressive subtypes. We therefore undertook mixture analyses (Agha & Ibrahim, 1984) to examine for any evidence of a bimodal distribution. First, summed SDRS scores for all 32 items were analysed, with tests examining for one versus two populations being non-significant. As the same test could not be undertaken for the Q-sort

**Table 5.** Validator variables

	Q-sort ( <i>n</i> = 228)			SDRS ( <i>n</i> = 228)			Newcastle ( <i>n</i> = 187)			DSM-IV ( <i>n</i> = 107)			Clinician ( <i>n</i> = 228)		
	Mel	N-Mel	Test	Mel	N-Mel	Test	Mel	N-Mel	Test	Mel	N-Mel	Test	Mel	N-Mel	Test
Mean age	48.5 (15.7)	39.8 (12.1)	<i>t</i> = -4.86***	48.4 (15.0)	40.0 (12.7)	<i>t</i> = -4.47***	52.3 (15.1)	40.9 (2.4)	<i>t</i> = -4.86***	46.1 (13.3)	40.1 (12.7)	<i>t</i> = -2.38*	48.5 (15.7)	39.4 (11.6)	<i>t</i> = -4.99***
Mean age of onset: first episode of depression	32.8 (13.9)	24.2 (12.3)	<i>t</i> = -4.23***	33.0 (13.9)	23.9 (12.0)	<i>t</i> = -4.62***	32.4 (15.3)	25.2 (12.2)	<i>t</i> = -3.11**	28.0 (12.5)	26.2 (12.1)	N.S.	32.4 (14.1)	23.8 (11.9)	<i>t</i> = -4.43***
Mean duration of current depressive episode in weeks	107.3 (157.5)	159.3 (295.8)	N.S.	100.6 (120.2)	164.0 (306.4)	N.S.	79.6 (81.0)	160.5 (289.1)	<i>t</i> = 2.75*	131.7 (196.6)	104.1 (98.4)	N.S.	87.1 (90.4)	174.0 (314.2)	<i>t</i> = 2.60*
Difficulty with parents (%)	25.9	50.8	$\chi^2 = 10.15^{**}$	32.3	48.4	$\chi^2 = 4.42^*$	30.8	45.9	N.S.	44.8	44.9	N.S.	22.7	54.1	$\chi^2 = 17.19^{***}$
Mean number of stressful life events (past 12 months)	1.6 (1.7)	2.3 (2.2)	( <i>n</i> = 187) <i>t</i> = 2.23*	1.6 (1.7)	2.3 (2.3)	( <i>n</i> = 187) <i>t</i> = 2.00*	1.3 (1.4)	2.3 (2.2)	( <i>n</i> = 187) <i>t</i> = 2.55*	2.3 (2.1)	2.4 (2.4)	N.S.	1.3 (1.5)	2.5 (2.3)	( <i>n</i> = 187) <i>t</i> = 4.08***
ECT (ever taken/currently taking) (%)	22.8	8.5	( <i>n</i> = 187) $\chi^2 = 7.29^{**}$	21.3	8.7	( <i>n</i> = 187) $\chi^2 = 5.82^*$	18.4	10.8	( <i>n</i> = 186) N.S.	17.2	4.1	( <i>n</i> = 107) $\chi^2 = 4.62^*$	27.7	4.9	( <i>n</i> = 187) $\chi^2 = 19.66^{***}$
Mean ECT effectiveness rating <sup>a</sup>	1.7 (0.9)	2.1 (0.7)	( <i>n</i> = 24) N.S.	1.8 (0.8)	2.0 (0.8)	( <i>n</i> = 24) N.S.	2.0 (0.8)	1.9 (0.8)	( <i>n</i> = 23) N.S.	2.0 (0.8)	1.5 (0.7)	( <i>n</i> = 12) N.S.	1.8 (0.8)	2.0 (0.9)	( <i>n</i> = 24) N.S.
Clinician-rated variables	(n = 188)			(n = 188)			(n = 187)			(n = 107)			(n = 188)		
Ongoing/pre-episode personality functioning limitations (%)	15.5	36.2	$\chi^2 = 8.17^*$	21.0	34.1	N.S.	7.7	35.8	$\chi^2 = 11.63^{***}$	27.6	36.7	N.S.	12.1	39.3	$\chi^2 = 15.18^{***}$
Mean overall functioning limitations score <sup>b</sup>	5.8 (1.3)	6.3 (1.4)	<i>t</i> = 2.13*	5.9 (1.3)	6.3 (1.4)	N.S.	5.6 (1.4)	6.3 (1.3)	<i>t</i> = 3.00**	6.0 (1.3)	6.2 (1.3)	N.S.	5.8 (1.3)	6.3 (1.3)	<i>t</i> = 2.45*

SDRS, Severity-based Depression Rating System; Mel, melancholic depression; N-Mel, non-melancholic depression; ECT, electro-convulsive therapy.

Note: *n* are indicated where analyses were based on subsamples of patients.

<sup>a</sup> Effectiveness ratings: 1 = not at all effective, 2 = somewhat effective, 3 = very effective.

<sup>b</sup> Higher scores indicate more severe dysfunction.

\* Significant at the *p* < 0.05 level (two-tailed). \*\* Significant at the *p* < 0.01 level (two-tailed). \*\*\* Significant at the *p* < 0.001 level (two-tailed).

scores because of the nature of the data (scores sum to zero), five items were selected on the basis that they produced a mean score difference of 1.0 or more between clinician-diagnosed melancholic and non-melancholic patients (see Table 1) in terms of Q-sort grid placement, and were most weighted to a clinical diagnosis of 'melancholia'. Neither Q-sort nor SDRS scores produced a bimodal distribution for this item set.

## Discussion

We sought to develop a self-report measure of melancholia, and by testing the same item set with contrasting prototypic and severity-based rating strategies, sought to quantify the extent to which the items themselves and/or the methodological strategy might influence differentiation. McKeown & Thomas (1988) noted that the Q-method 'retains a somewhat fugitive status within the larger scientific community'. Our results suggested a slight advantage to the prototypic approach, as the Q-sort strategy effectively (through logistical regression analyses) identified 81.6% of the total sample as having a 'melancholic' or 'non-melancholic' depression when compared to Clinician diagnosis. In light of the statistically comparable AUC values for both measures derived from ROC analyses, we tentatively position the Q-sort strategy as a potential alternative to the SDRS as a screening tool for melancholic depression because of its higher specificity and negative predictive values when using optimal cut-off scores based on the total sample. We suggest that by testing such an approach, and quantifying comparative classification success of prototypic and severity-based measures, we allow other researchers the opportunity to consider whether a prototypic approach may have utility to a similar or independent inquiry. It may be that, in some inquiries, rating severity of features is more important than rating characteristic features, whereas for other inquiries (particularly in differentiating clinical subtypes) the converse may hold. The Q-sort strategy effectively 'forces' subjects to rank items according to a normal distribution, thus overcoming any tendency to rate at the extreme ends of any dimensional scale, and when such ratings can reflect contrasting minimizing and maximizing biases.

In any sample of depressed patients, some will evidence 'clear-cut' characteristic depressive patterns whereas others will have less distinctive patterns. We elected to focus our analyses on all subjects rather than the more diagnostically 'confident' or clear-cut subjects, as the latter approach theoretically risks optimizing classification rates, and with such subjects not likely to represent those presenting to primary

and secondary care facilities. Our overall sample comprised patients referred to a tertiary consultative service, with many having diagnostically unclear conditions and generating low 'confidence' ratings. In light of this context issue, the high differentiation achieved by the Q-sort strategy in this sample is noteworthy. As our severity-based (SDRS) measure was almost as successful, it is likely that high differentiation emerged more from the item set rather than from the contrasting strategy. As the Q-sort was completed prior to the SDRS, many might have been cued by the Q-sort procedure preceding the SDRS presentation, artificially assisting SDRS differentiation, and arguing for a counterbalanced approach in replication studies.

Examined against several potential validators, additional support for the Q-sort solution emerged. Those assigned by the Q-sort strategy to a melancholic class had a profile consistent with ascriptions of 'melancholia': that it appears at an older age, has a more 'endogenous' background rather than reflect distal and proximal antecedent psychosocial factors, and is more likely to require ECT.

Q-sort assignment (though showing some agreement with clinician, SDRS and Newcastle assignments) was at considerable variance with DSM-IV assignment, whereas only one validation variable (i.e. receipt of ECT) was significant in relation to DSM-IV assignment. Thus, Q-sort replication and extension (e.g. aetiological) studies should focus on both intrinsic utility and comparative utility with DSM-IV decision rules.

One key limitation is noted. By using a clinician-based diagnostic decision as the reference criterion, there is a risk of circularity, in clinically weighting the diagnosis of melancholia to certain items, and also selecting similar items for the Q-sort, so creating associations. However, the selected Q-sort items respected historical emphases in describing melancholia. In addition, as demonstrated in Table 1, many of the selected items emerged as minimally differentiating or non-differentiating, arguing against a simple circularity process. Finally, high (blinded) inter-rater agreement in judging whether a patient had a melancholic or non-melancholic depressive disorder argues against idiosyncratic diagnostic subtyping.

Although we asked patients to ignore medication side-effects, we accept that any such impact cannot be reliably ensured. However, if operative, we would anticipate that it would have impacted similarly on both rating strategies and therefore been largely 'controlled'.

Although we sought to measure 'melancholia' rather than establish its status as a categorical subtype, analyses failed to find any evidence of a bimodal



distribution of scores, which would support a categorical binary model. However, no symptom item showed absolute specificity to either diagnostic subtype, with most (whether weighted to melancholic or non-melancholic depression) returning minor mean differences at best. Thus, and as indicated in the Introduction, most depressive symptoms are imprecise and lack clear specificity. Exceptions might include symptoms that tap categorical constructs (e.g. psychotic symptoms or psychomotor signs).

In our previous studies of melancholia (Parker & Hadzi-Pavlovic, 1996), latent class analyses of CORE item probabilities established low differentiation of symptoms at best, whereas 'signs' of psychomotor disturbance had low latent class probabilities in the non-melancholic class and distinct probabilities in the non-melancholic class. This allowed us (in conjunction with mixture analyses rejecting a unimodal distribution) to conclude that observable psychomotor disturbance went a considerable way to meeting 'necessary and sufficient' criteria for the definition of melancholia. Thus, bimodality is likely to depend on the presence of items that are specific to one class and as we had no such variable (e.g. psychotic symptoms) in our item set, such a distribution could not be demonstrated.

We can reconcile results from these two differing study approaches to suggest that a percentage of those with 'true melancholia' are likely to have psychotic features and another percentage are likely to have overt psychomotor disturbance. If melancholia is defined by the presence of one or both of those features, then categorical distinction is likely to be able to be made to those with residual depressive conditions, and a bimodal distribution demonstrated on quantifying such constituent features. However, if melancholia is alternatively defined merely by a set of symptoms that are over- or under-represented in those with melancholia (as indicated here and in general), the absence of clear-cut symptom specificity will prevent bimodality being demonstrated. Our study was designed to develop a symptom measure of melancholia, not to test the hypothesis that melancholia is a categorical entity.

Our measure could be reduced from its current set of 32 items, preserving those most specific to both melancholia and non-melancholic depression. However, we would argue that the initial step would encompass replication studies preserving the full set for testing across differing rating instructions, particularly in evaluating their prototypic status as against severity-based strategies. Refinement of the item set, and decisions as to whether a Q-sort strategy is the optimal strategy, might then best occur following analysis of multiple data sets. We do note, however, that item

reduction does not necessarily present a key objective for the Q-sort strategy, as any derived algorithms rely on item placements within the grid as opposed to the specific number of items selected. We conclude that the present study has advanced the objective of producing a symptom-based strategy for distinguishing melancholic and non-melancholic depression. Future research will seek to clarify and refine item-sets adopting the Q-sort strategy or any strategy that seeks to measure 'melancholia' on the basis of the comparative relevance of symptoms descriptors, and in particular in determining whether the prototypical nature of symptoms, as opposed to symptom severity, is the more salient domain to differentiate any melancholic depressive subtype.

### Acknowledgements

This work was supported by Program Grant 510135 from the National Health and Medical Research Council of Australia, an Infrastructure Grant from NSW Health and by the Heine Foundation.

### Declaration of Interest

None.

### References

- Agha M, Ibrahim MT (1984). Algorithm AS 203: maximum likelihood estimation of mixtures of distributions. *Applied Statistics* **33**, 327–332.
- Bech P (2002). The Bech–Rafaelsen Melancholia Scale (MES) in clinical trials of therapies in depressive disorders: a 20-year review of its use as outcome measure. *Acta Psychiatrica Scandinavica* **106**, 252–264.
- Bech P, Rafaelsen OJ (1980). The use of rating scales exemplified by a comparison of the Hamilton and the Bech–Rafaelsen Melancholia Scale. *Acta Psychiatrica Scandinavica* **62** (Suppl. 285), 128–132.
- Block J (1961). *The Q-sort Method in Personality Assessment and Psychiatric Research*. Charles C. Thomas: Springfield, IL.
- Carney MW, Roth M, Garside RF (1965). The diagnosis of depressive syndromes and the prediction of E.C.T. response. *British Journal of Psychiatry* **111**, 659–674.
- Demyttenaere K, De Fruyt J (2003). Getting what you ask for: on the selectivity of depression rating scales. *Psychotherapy and Psychosomatics* **72**, 61–70.
- Jones EE (1985). Manual for the Psychotherapy Process Q-sort. Unpublished manuscript, University of California, Berkeley.
- Kendell RE (1989). Clinical validity. *Psychological Medicine* **19**, 45–55.
- Kraemer HC (1992). *Evaluating Medical Tests: Objective and Quantitative Guidelines*. Sage: Newbury Park, CA.
- Leventhal AM, Rehm LP (2005). The empirical status of melancholia: implications for psychology. *Clinical Psychology Review* **25**, 25–44.

- McKeown BF, Thomas DB** (1988). *Q Methodology*. Sage: Newbury Park, CA.
- Nelson JC, Charney DS** (1981). The symptoms of major depressive illness. *American Journal of Psychiatry* **138**, 1–13.
- Parker G, Hadzi-Pavlovic D** (1996). *Melancholia: A Disorder of Movement and Mood*. Cambridge University Press: New York.
- Parker G, Hilton T, Bains J, Hadzi-Pavlovic D** (2002). Cognitive-based measures screening for depression in the medically ill: the DMI-10 and DMI-18. *Acta Psychiatrica Scandinavica* **105**, 419–426.
- Pilowsky I, Levine S, Boulton DM** (1969). The classification of depression by numerical taxonomy. *British Journal of Psychiatry* **115**, 937–945.
- Rush AJ, Hiser W, Giles DE** (1987). A comparison of self-reported versus clinician-related symptoms in depression. *Journal of Clinical Psychiatry* **48**, 246–248.
- Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB** (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS) Clinician Rating (QIDS-C) and Self-Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry* **54**, 573–583.
- Rush AJ, Weissenburger JE** (1994). Melancholic symptom features and DSM-IV. *American Journal of Psychiatry* **151**, 489–498.
- Smolka M, Stieglitz RD** (1999). On the validity of the Bech–Rafaelsen Melancholia Scale (BRMS). *Journal of Affective Disorders* **54**, 119–128.
- Spitzer RL** (1983). Psychiatric diagnosis: are clinicians still necessary? *Comprehensive Psychiatry* **24**, 399–411.
- Taylor MA, Fink, M** (2006). *Melancholia: The Diagnosis, Pathophysiology, and Treatment of Depressive Illness*. Cambridge University Press: New York.
- Watts S, Stenner P** (2005). Doing Q methodology: theory, method and interpretation. *Qualitative Research in Psychology* **2**, 67–91.
- Zimmerman M, Coryell W** (1987). The Inventory to Diagnose Depression (IDD): a self-report scale to diagnose major depressive disorder. *Journal of Consulting and Clinical Psychology* **55**, 55–59.