Journal of Affective Disorders 144 (2013) 199–207

Contents lists available at SciVerse ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research report

Discriminating melancholic and non-melancholic depression by prototypic clinical features

Gordon Parker [a,b,*], Stacey McCraw [a,b], Bianca Blanch [a,b], Dusan Hadzi-Pavlovic [a,b], Howe Synnott [b], Anne-Marie Rees [a,b]

[a] School of Psychiatry, University of New South Wales, Sydney, NSW, Australia
[b] Black Dog Institute, Sydney, Australia

ARTICLE INFO

Article history:
Received 5 January 2012
Received in revised form
27 June 2012
Accepted 27 June 2012
Available online 4 August 2012

Keywords:
Depression
Melancholia
Depression measurement
Discriminant index
Sydney Melancholia Prototype Index.

ABSTRACT

Background: Melancholia is positioned as either a more severe expression of clinical depression or as a separate entity. Support for the latter view emerges from differential causal factors and treatment responsiveness but has not been convincingly demonstrated in terms of differential clinical features. We pursue its prototypic clinical pattern to determine if this advances its delineation.

Methods: We developed a 24-item measure (now termed the Sydney Melancholia Prototype Index or SMPI) comprising 12 melancholic and 12 non-melancholic prototypic features (both symptoms and illness correlates). In this evaluative study, 278 patients referred for tertiary level assessment at a specialized mood disorders clinic completed the self-report SMPI as well as a depression severity measure and a comprehensive assessment schedule before clinical interview, while assessing clinicians completed a clinician version of the SMPI items following their interview. The independent variable (diagnostic gold standard) was the clinician's judgment of a melancholic versus non-melancholic depressive episode. Discriminative performance was evaluated by Receiver Operating Characteristics (ROC) analysis of four strategies for operationalising the SMPI self-report and SMPI clinician measures, and with the former strategies compared to ROC analysis of the depression severity measure. The external validity of the optimally discriminating scores on each measure was tested against a range of clinical variables.

Result: Comparison of the two self-report measures established that the SMPI provided greater discrimination than the depression severity measure, while comparison of the self-report and clinician-rated SMPI measures established the latter as more discriminating of clinically diagnosed melancholic or non-melancholic depression. ROC analyses favoured self-report SMPI distinction of melancholic from non-melancholic depression being most optimally calculated by a 'difference' score of at least four or more melancholic than non-melancholic items being affirmed (sensitivity of 0.69, specificity of 0.77). For the clinician-rated SMPI measure, ROC analyses confirmed the same optimal difference score of four or more as highly discriminating of melancholic and non-melancholic depression (sensitivity of 0.84, specificity of 0.92). As the difference score had positive predictive values of 0.90 and 0.70 (for the respective clinician-rated and self-report SMPI forms) and respective negative predictive values of 0.88 and 0.70, we conclude that the clinician-rated version had superior discrimination than the self-report version. External validating data quantified the self-rated and clinician-rated Index-assigned non-melancholic patients having a higher prevalence of anxiety disorders, a higher number of current and lifetime stressors, as well as elevated scores on several personality styles that are viewed as predisposing to and shaping such non-melancholic disorders.

Limitations: Assigned melancholic and non-melancholic diagnoses were determined by clinician judgement, risking a circularity bias across diagnostic assignment and clinical weighting of melancholic and non-melancholic features. The robustness of the Index requires testing in primary and secondary levels of care settings.

Conclusions: The clinician-rated SMPI differentiated melancholic and non-melancholic depressed subjects at a higher level of confidence than the self-report SMPI, and with a highly acceptable level of discrimination. The measure is recommended for further testing of its intrinsic and applied properties.

* Corresponding author at: Black Dog Institute, Prince of Wales Hospital, Randwick 2031, Sydney, Australia. Tel.: +61 2 9382 4372; fax: +61 2 9382 4343.
E-mail address: g.parker@unsw.edu.au (G. Parker).

0165-0327/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jad.2012.06.042

1. Introduction

There has been a longstanding view positioning melancholia as a distinct 'type' of depression that appears more quintessentially biological, and which has been variably termed 'endogenous', 'endogenomorphic', 'autonomous', 'vital', 'Type A' as well as 'melancholic' depression (Jackson, 1986; Parker et al., 2010; Parker and Hadzi-Pavlovic, 1996a; Taylor and Fink, 2006). The arguments in favour of its distinct status have included (Parker et al., 1996, 2010) a somewhat distinctive pattern of symptoms and signs, a greater relevance of genetic and other biological – as against psychosocial – determinants, concomitant evidence of biological dysfunction particularly involving the hypothalamic–pituitary–adrenal axis, a differentially stronger response to physical treatments such as antidepressant drugs and electroconvulsive therapy than to psychotherapy, and a low placebo response rate.

The longstanding binary view positioned such a depressive 'type' as distinct from a second 'type' – variably termed 'neurotic' or 'reactive' depression in terms of clinical symptoms and preferential causes. Despite some consistency in the clinical ('endogeneity') symptoms long listed as having some specificity to melancholia, the advent and application of differing multivariate analytic approaches in the 1950's – whether factor, cluster and (later) latent class analysis (Parker and Hadzi-Pavlovic, 1996a) – failed to deliver support for a clear-cut symptom-defined binary solution.

Reasons for failing to so differentiate melancholia would include it actually not being a distinctive depressive type and simply being a more severe expression of depression—essentially the unitary or dimensional view. Alternatively, it could be that melancholia is a differing 'type' but lacking a pristine clinical boundary so disallowing clear clinical delineation and/or that its putative clinical symptoms and signs are limited in terms of their specificity and capacity to define melancholia. For example, and as we have quantified (Parker et al., 1996), none of the historically-weighted endogeneity symptoms show absolute specificity and, at best, show modest differential prevalences. Even if a symptom demonstrates discriminatory potential, identifying how it is best operationalised and measured is rarely straightforward. If not absolute, deciding whether to impose a cut-off for its 'presence' along dimensions of severity, persistence or some other parameter is problematic. Further, age, gender and duration of episode may impact on symptom ratings, while response biases (e.g., excessive subjective weighting versus denial and minimisation) influence self-reporting—just as assessment by external observers can be influenced by rating biases.

Finally, if melancholic depression is a 'circuit disorder' involving disruption of neurocircuits, then the actual site or dynamics of the disruption may account for certain symptoms (e.g., abulia, psychomotor agitation) being distinctive in some individuals and minimal or even absent in others. Thus, even if melancholia is a discrete condition, its symptom markers are limited by multiple factors that must confound any analytic study seeking to delineate it simply on the basis of symptoms with any precision.

Historical approaches to defining and classifying melancholia over recent decades have involved relatively few strategies. First, and most commonly, limiting definition to a prescribed number of symptoms (as occurs in DSM-IV). Second, melding clinical symptoms with non-symptom correlates of melancholia. The latter approach has only a few examples. One was the Newcastle Index (Carney et al., 1965) which weighted items such as 'adequate personality', 'no adequate psychogenesis' and previous episodes in addition to symptoms. Another was the DSM-III-R classification of melancholia which included items such as absence of any pre-morbid personality disturbance, previous episodes with good recovery and previous good response to somatic therapies in addition to symptoms.

Narrower strategies have been evaluated. First, weighting and measuring signs of psychomotor disturbance (PMD), with the view that such observable signs are surface markers of underlying neuropathological processes in melancholia, a model reflecting PMD's longstanding position as a central marker of melancholia (Berrios, 1988). Following on Widlöcher's (1983) development of a refined measure, we developed the observer-rated CORE measure (Parker and Hadzi-Pavlovic, 1996b) of PMD—with that measure so named to capture its reference to 'core' signs of melancholia. Limitations to rating signs validly include the reality that not all patients present at the nadir or depths of their depressive episode and that the motor signs of PMD are seemingly less overt or severe in younger melancholic patients.

In the last few years, we have favoured diagnostic measurement melding clinical features and non-symptom correlates, and offer several reasons. First, the approach concedes limitations (just detailed) to relying on any symptom set alone. Second, it reflects a number of the longstanding prototypic ascriptions to the concept of melancholia—with even its synonym 'endogenous depression' proceeding beyond symptoms. Third, it is consistent with the approach to defining many medical conditions (e.g., Parkinson's disease) and where diagnosis is based on a range of antecedent and course of illness factors in addition to symptoms. Fourth, we have already demonstrated (Parker et al., 2010) that adding course of illness and context variables to refined symptoms actually improves delineation of melancholic and non-melancholic depression made by symptom definition alone—and in that report made an analogy to navigational strategies that rely on multiple reference points to improve precision. Further, it acknowledges the likely reality that melancholia is 'fuzzy', and suggests that definition might better be weighted to prototypic delineation rather than to seeking absolute definition.

We therefore developed (Parker et al., 2012) the SERDEX measure (SElf-Report of Depressive EXperiences) which lists 12 items weighted to melancholic depression in a left-hand column and 12 items weighted to the non-melancholic depressive conditions in a right-hand column. Individuals are invited to tick any item from either column that they regard as 'characteristic' in terms of their depressive experience, whether (dependent on the clinical or study objective) experienced currently or over time. The listed items assess symptoms historically favoured as most differentiating of melancholic and non-melancholic depression, but also assess premorbid interpersonal functioning, distal and proximal stressors, the context and impact of proximal stressors on inducing and maintaining the depression, and trait emotional dysregulation levels. Each item was selected and often progressively refined in its definition by considering its utility in previous studies undertaken by our research group over the last twenty years (e.g., Parker and Hadzi-Pavlovic, 1996a; Parker et al., 2010) and with all having been tested empirically to quantify their differentiating potential. For example, while early morning wakening is commonly listed as a symptom of 'endogenous' or 'melancholic' depression, we have never quantified it as having distinctive differentiation across melancholic and non-melancholic depression and it was therefore not included. After ticking relevant *items*, respondents are then requested to judge whether their 'profile' or clinical *prototype* is best captured by Description A (left-hand column descriptors), Description B (right-hand column), is somewhat closer to A than to B, is somewhat closer to B than to A, or is an equal mix or A and B descriptors—with this second 'prototypic' measurement component seeking to determine overall 'pattern' correspondence to melancholic or non-melancholic depression. For the present study we developed an equivalent clinician-rated version of the measure.

We reported the properties of the initial self-report measure in an earlier paper (Parker et al., 2012) with that development study involving a sample of 141 unipolar depressed patients assessed at

our tertiary referral Black Dog Institute, and demonstrated that non-symptom variables were superior to symptoms in differentiating melancholic and non-melancholic depression. In this paper, we analyse diagnostic differentiation and validation data from a larger sample and, in addition, examine the utility of a clinician-rated version of the measure.

2. Methods

As for the previous pilot study (Parker et al., 2012), the current sample was recruited entirely from the Depression Clinic at the Black Dog Institute, a tertiary service providing diagnostic and management advice to referring general practitioners and psychiatrists. While weighted to those with more severe and/or treatment-resistant mood disorders, not all patients were assessed at the nadir of their episode (including some being in partial remission) as a consequence of Clinic waiting time. The Human Research Ethics Committee for the University of New South Wales approved the study protocol. All participants were provided with a description of the study prior to clinical assessment and written informed consent was then obtained.

For the current analyses, we report on a larger sample of 278 patients. All were required to score symptom items in terms of their current episode (''when at your worst'') and whether the symptom was a characteristic feature of their episode or not, with all questionnaires being completed prior to clinical assessment. Sample recruitment occurred from May 2009 to December 2010. While we did not formally record the non-participation rate, we estimate it as low (in the order of ten per cent) and generally due to poor English skills or an incapacitating mood state compromising some patients' capacities to complete questionnaires.

For all patients, the Clinic's six clinicians were required to make a judgement as to whether they diagnosed the patient as having a current melancholic or non-melancholic depressive episode, and to rate their level of confidence (1–5) in the diagnostic allocation to melancholic or non-melancholic status, but, in this study we did not exclude any patients on the basis of the confidence score. Such clinician judgements were derived from their clinical interview – which focuses on depressive symptom patterns, examines family history, developmental factors, personality profile and previous response to any drug and non-drug treatments – and sometimes involves an interview of corroborative witnesses (e.g., family members, referring health practitioners). Intake clinicians also completed a clinician-rated version of the SERDEX measure for the current depressive episode and again required affirming any symptom that was characteristic when the patient was at their worst.

While we had varying numbers of recruited patients complete differing assessment components, this report (and analyses) were restricted to 278 patients for whom we had complete data sets in relation to our principal measures (i.e., SERDEX items, depression severity and Mood Assessment Program). Of the sample of 278 (51.4% female, mean age 41.2 years), 43.5% were clinically diagnosed with a melancholic depression and were non-significantly older than the non-melancholic subjects (42.6 vs 40.2, $t=1.5$, $p=0.14$). Of the 278 subjects, the rating clinicians failed to complete the five-point prototype scale at the bottom of the index for 11 subjects, so that analyses of that variable involved 267 of the 278 subjects.

2.1. Materials

In addition to completing the SERDEX form, all subjects completed a self-report Severity of Depressive Symptoms (SDS) form (Parker et al., 2010) comprising 32 clinical symptoms of depression rated on 0–3 scales, allowing respective rating options of 'not at all', 'somewhat', 'very' and 'extremely' characteristic. The 32 symptoms captured ones historically weighted to both melancholic and non-melancholic depression. Comparative analyses of those two key study measures would allow the utility of each approach (i.e., symptoms only versus symptoms plus illness correlates) to be compared. Sample members also completed the Mood Assessment Program or MAP (Parker et al., 2008), a computerised measure providing current and historical data on depression and anxiety, current global functioning, lifetime anxiety disorders, the Temperament and Personality or T&P personality profile (Parker et al., 2006) and stressor severity measures (both lifetime and preceding 12 months), as well as recording background and lifetime treatment data – with such variables being analysed in the validation component of this study. State depression severity was measured by both the DMI-10 (Depression in the Medially Ill) measure (Parker et al., 2002) – which provides a 0–30 severity score – and The Quick Inventory of Depressive Symptomatology (QIDS-SR16; Rush et al., 2003) – which provides a 0–27 severity score.

3. Results

3.1. Item discrimination across the SERDEX measures

Table 1 reports prevalence rates for all 24 items in the clinically diagnosed melancholic and non-melancholic subgroups. If items are truly more prevalent in melancholia, we would anticipate that all 12 descriptor A items would return significantly higher prevalence rates in the melancholic than the non-melancholic sub-set—as confirmed (examining chi square data) for 6 of the items in the self-report measure and all 12 in the clinician-rated measure. Conversely, we would expect that all 12 descriptor B items would return higher prevalence rates in the non-melancholic than melancholic sub-set—confirmed for 8 self-report and 10 clinician-rated items. Third, if items had high specificity we would expect distinctive differentiation in prevalence of items across the melancholic and non-melancholic subjects. Inspection of Table 1 data fails to identify any item selectively affirmed by one diagnostic sub-set only, and that item differentiation across melancholic and non-melancholic sub-sets was slight to moderate only, but—and of importance - far more distinctive for the clinician-rated than for the self-report measure.

Table 1 reports prevalence data for each item and with any over-representation of putative melancholic items being affirmed by the subjects and by the assessing clinician in the melancholic subjects (and of putative non-melancholic items in the non-melancholic subjects) quantified by odds ratios. Across the self-report and clinician-rated SERDEX forms the most discriminating Descriptor A (melancholia) items were anergia, episodes being more severe than expected from circumstances, episodes 'coming out of the blue' and anticipatory anhedonia. For the non-melancholic Descriptor B set, the most discriminating items across both rating approaches were whether or not the episode severity was 'explainable', anticipatory anhedonia was absent, there was an explanatory cause, mood reactivity, and concentration impairment reflecting worry or distracting thoughts. In essence, the most discriminating items were consistently identified across self-report and clinician-rated assessment and such items were equally likely to be symptom or non-symptom variables.

3.2. Analyses of prototypic assignments

Table 2 records how the clinically diagnosed melancholic and non-melancholic members scored the single five-point self-report

Table 1
Items endorsed by sample member for self-report and clinician rated SERDEX measures for clinically diagnosed melancholic (Mel) and non-melancholic (Non-Mel) patients.

Description A	Self-report measure ($n=278$)				Clinician measure ($n=278$)			
	Mel (%)	Non-mel (%)	χ^2	OR	Mel (%)	Non-mel (%)	χ^2	OR
1. Low energy and extremely hard to get out of bed and get going	82.6	61.8	14.4***	2.9***	89.3	43.3	62.1***	10.9***
2. Depressed mood completely prevents getting any real pleasure in life, and normally pleasing or humorous things won't lift mood	81.0	70.7	3.9*	1.8*	82.6	32.5	69.3***	9.9***
3. Mood and energy levels are worse in the mornings	67.8	49.7	9.2*	2.1*	62.8	24.2	42.1***	5.3***
4. Completely lose interest in things, including hobbies and activities that would usually be enjoyed when not depressed	81.0	71.3	3.4	1.7	87.6	39.5	66.2***	10.8***
5. Cannot look forward to anything in life	78.5	56.1	15.3***	2.9***	79.3	33.1	58.6***	7.8***
6. In walking and talking, distinctly physically slowed, at times almost feeling 'paralysed' or as if walking through sand	46.3	31.8	6.0*	1.8*	54.5	10.8	62.4***	9.9***
7. Concentration is distinctly affected	84.3	75.8	3.0	1.7	87.6	61.1	24.1***	4.5***
8. Tend to lose weight when depressed (and before any antidepressant or other drugs are commenced)	28.9	25.5	0.4	1.2	43.8	15.3	27.7***	4.3***
9. The severity of depressive episodes appears far worse than would be expected given the circumstances that may precede them or appear to cause them	66.1	42.7	15.1***	2.6***	73.6	15.3	96.2***	14.4***
10. Early years were no more difficult – when compared to most people – in terms of having any major difficulties with parents or bullying	53.7	36.3	8.4*	2.0*	52.9	27.4	18.8***	3.0***
11. When not depressed relationships and work performance are generally good	77.7	68.8	2.7	1.6	72.7	38.2	32.7***	4.3***
12. Depressions can sometimes come 'out of the blue' without any particularly clear reason	74.4	51.6	15.0***	2.7***	70.2	24.8	57.0***	7.1***
Description B	**Self-report measure**				**Clinician measure**			
	Mel (%)	Non-mel (%)	χ^2	OR	Mel (%)	Non-mel (%)	χ^2	OR
1.Even when depression is severe, can generally look forward to something really nice coming up	6.6	21.7	12.1**	3.9**	5.8	40.1	42.8***	10.9***
2. Become distinctly more irritable and/or angry when depressed	61.2	63.7	0.2	1.1	49.6	56.7	1.4	1.3
3. Even when depression is severe, can generally be cheered up when people are really supportive	23.1	23.6	0.0	1.0	9.1	48.4	49.1***	9.4***
4. Mood lifts (even if temporarily) and can get some temporary relief when something nice happens	39.7	55.4	6.8*	1.9*	20.7	64.3	52.6***	6.9***
5. If concentration is affected during a depressive episode, it is usually because of worrying too much and having lots of distracting thoughts	57.9	72.6	6.7*	1.9*	19.0	59.2	45.5***	6.2***
6. Often get (non-medication related) food cravings and/or increased appetite when depressed	32.2	37.6	0.9	1.3	24.8	31.8	1.7	1.4
7. Views self as generally more inclined than most people to become emotional about things (regardless of whether depressed or not)	38.8	53.5	5.9*	1.8*	22.3	45.2	15.7***	2.9***
8. Every time depression develops, some cause that explains the depression is apparent	25.6	43.3	9.3*	2.2*	9.9	53.5	57.2***	10.5***
9. The severity of depressions can be explained by the type of stressful events that precede them and their impact with personality style	20.7	45.9	19.1***	3.3***	6.6	57.3	77.0***	19.0***
10. Even when not depressed, tends to have some difficulties in dealing with my partner, family and other relationships	28.1	39.5	3.9*	1.7*	17.4	43.3	21.2***	3.6***
11. Even when not depressed, tend to worry more than most people, particularly when under stress	52.1	58.0	1.0	1.3	25.6	59.9	32.4***	4.3***
12. In childhood and adolescence, experienced more stressful events and major difficulties with parents and others than most people experience	28.1	52.2	16.4***	2.8***	28.1	49.7	13.2***	2.5**

Items are slightly modified from actual self-report and clinician-rated forms.

* $p<0.05$.
** $p<0.01$.
*** $p<0.001$.

Table 2
Overall descriptor pattern endorsed by clinically diagnosed melancholic and non-melancholic patients according to the single-item self-rated ($n=278$) and clinician-rated ($n=267$) prototypic measure.

SERDEX ratings by clinically diagnosed melancholic and non-melancholic patients	1 Description A best matches my experience of depression	2 Description A is somewhat closer to my experience than B	3 My depression has equal features of Descriptions A and B	4 Description B is somewhat closer to my experience than A	5 Description B best matches my experience of depression
Melancholic patients					
Self-rated SERDEX	42 (34.7%)	49 (40.5%)	19 (15.7%)	11 (9.1%)	0 (0.0%)
Clinician-rated SERDEX	55 (47.8%)	48 (41.7%)	8 (7.0%)	3 (2.6%)	1 (0.9%)
Non-melancholic patients					
Self-rated SERDEX	17 (10.8%)	32 (20.4%)	63 (40.1%)	30 (19.1%)	15 (9.6%)
Clinician-rated SERDEX	3 (2.0%)	10 (6.6%)	37 (24.3%)	64 (42.1%)	38 (25.0%)

Examining for differential distribution for those clinically judged as either having a melancholic or non-melancholic depression, the clinician-rated SERDEX generated a chi square of 179.2 ($p<0.001$) and the self-rated SERDEX a chi square of 57.9 ($p<0.001$).

prototypic measure or were so rated by the clinician in regard to their overall clinical profile as matching Description A (prototypic melancholia), Description B (prototypic non-melancholic depression) or a mix. Data for the self-report version quantified that 75.2% of the clinically-diagnosed melancholic patients favoured Description A ('best' or 'closer') and only 9.1% favoured Description B as 'best' or 'closer'. For the clinician-rated measure, discrimination was even more distinctive (i.e., 89.5% vs. 3.5%,

respectively). Turning to the clinically-diagnosed non-melancholic patients, there was far less specificity on the self-report measure (28.7% favoured Description B as 'best' or 'closer' and 31.2% favoured Description A) but with superior rates (of 67.1% versus 8.6%, respectively) for the clinician-rated prototype measure. Thus, on the self-report measure, Description A was weighted by the melancholic patients and showed distinct skewing, while the non-melancholic patients showed a spread across the five options and with their pattern more approximating to a normal distribution rather than being loculated to Description B. By contrast, the clinician-rated measure was skewed to the putative diagnostic category.

We examined agreement between self-report and clinician-rated forms, and quantified weighted kappas of 0.21 for the melancholic and 0.09 for the non-melancholic subjects, establishing a lack of agreement between self-report and clinician-rated assignment to one of the five prototypic categories, and suggesting limitations to one or both measures.

3.3. Examining the comparative capacity of the derived self-report SERDEX and the SDS symptom severity measures to differentiate clinically defined melancholic and non-melancholic depression

We undertook a series of ROC analyses examining the capacity of a (i) 'total melancholia' score (i.e., number of the 12 Description A items affirmed), (ii) 'total non-melancholia' score (i.e., number of the 12 Description B items affirmed), (iii) total 'difference' score (i.e., subtracting the total number of affirmed Description B from affirmed Description A items) and the (iv) 'prototype' assignment chosen by the patient (i.e., A, A > B, A = B, B > A and B options). As graphed and quantified in Fig. 1, the least discriminating measures were the 'total melancholia' score (AUC = 0.72, CI = 0.66–0.78, $p < 0.001$) and 'total non-melancholia' score (AUC = 0.69, CI = 0.63–0.75, $p < 0.001$) and with both comparable in their discrimination (CI = −0.04–0.12, $p = 0.41$). The prototype score was only slightly more discriminating (AUC = 0.74, CI = 0.69–0.80; $p < 0.001$) than these two strategies. A total 'difference' score of 4 or more melancholic than non-melancholic items was, however, the most discriminating in terms of its AUC (0.76), being formally superior to the total non-melancholic score (CI = −0.11–0.03, $p < 0.001$) and non-significantly superior to both the total melancholia score (CI = −0.08–0.00, $p < 0.06$) and the prototype score (CI = −0.01–0.05, $p < 0.25$). The total 'difference' score of 4 or more was therefore judged as the optimal discriminating strategy, and quantified as having a sensitivity of 0.69, specificity of 0.77, positive predictive value of 0.70 and negative predictive value of 0.77.

We then undertook a similar analysis of total SDS severity scores, identified a total score of 54 or more as having the greatest discrimination of clinically diagnosed melancholic and non-melancholic patients (AUC = 0.68, CI = 0.62–0.75, $p < 0.001$) and thus far less discriminating than all four applications of the SERDEX measure. That SDS cut-off score had a sensitivity of 0.61, specificity of 0.65, positive predictive value of 0.57 and negative predictive value of 0.69. Fig. 1 plots the ROC curves for the four self-report SERDEX applications and for the total SDS score, demonstrating the superiority all four SERDEX approaches to the SDS measure and, as noted in the previous paragraph, that of the four SERDEX applications, the greatest differentiation was evident in the SERDEX 'difference' score of 4 or more.

3.4. Examining the comparative capacity of derived clinician-rated SERDEX measures to differentiate clinically defined melancholic and non-melancholic depression

The series of ROC curve analyses performed for the self-rated SERDEX were repeated for the clinician-rated SERDEX measure (see Fig. 2). The total number of melancholic items (AUC = 0.93, CI = 0.90–0.96, $p < 0.001$) and the difference score (AUC = 0.93, CI = 0.90–0.96, $p < 0.001$) were formally comparable in their discrimination (CI = 0.02–0.02, $p = 0.70$). The least discriminating

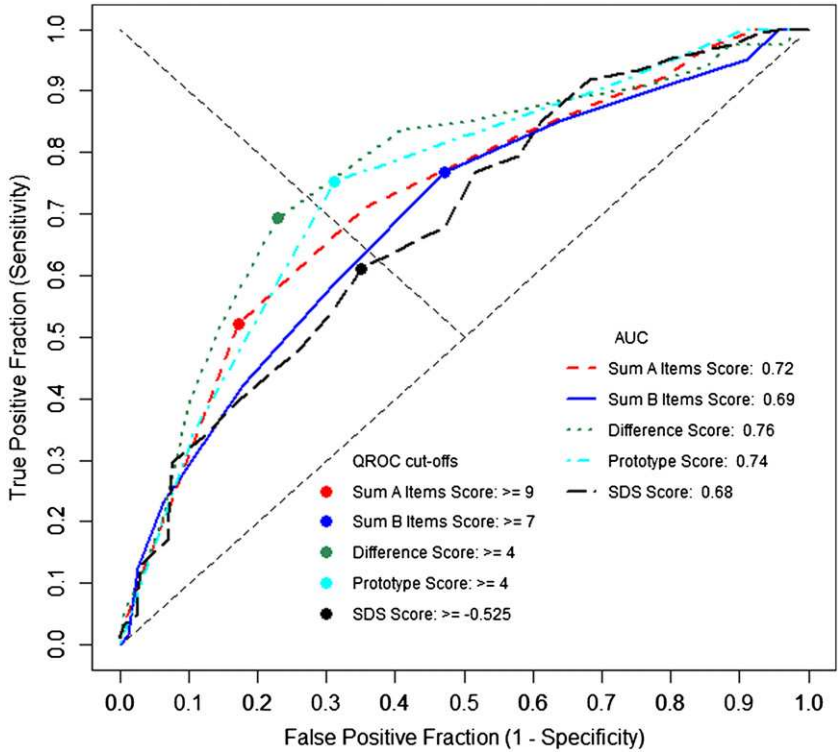

Fig. 1. ROC curve analysis for the sum of melancholic (A) items, non-melancholic (B) items, the difference (A)–(B) score, prototype score and SDS total score for the self-rated SERDEX measure.

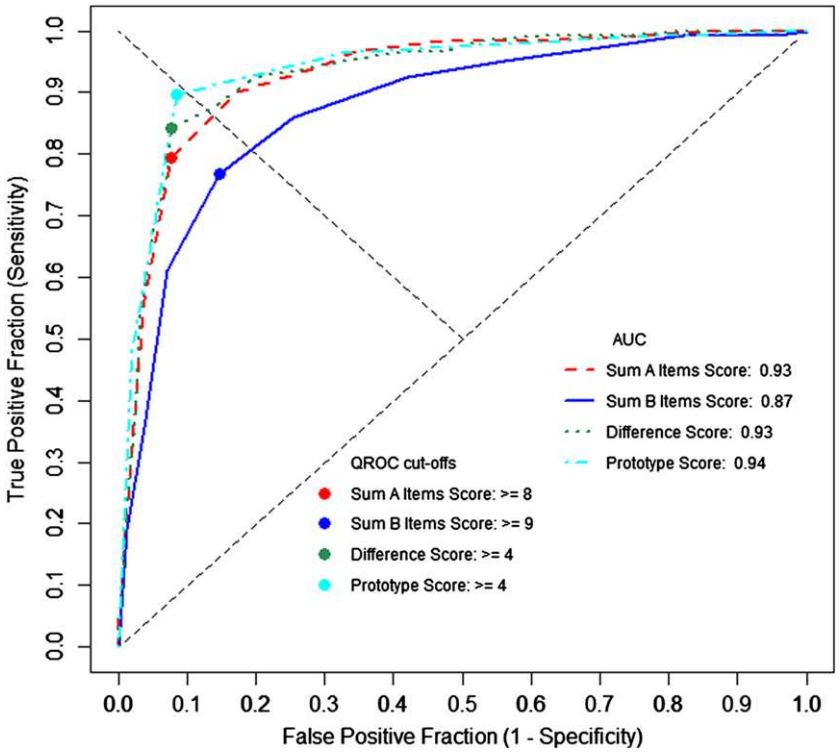

Fig. 2. ROC curve analysis for the sum of melancholic (A) items, non-melancholic (B) items, the difference (A)–(B) score, prototype score and SDS total score for the clinician-rated SERDEX measure.

measure was the total number of non-melancholic items affirmed (AUC=0.87). While the formally most discriminating measure was the prototype score (AUC=0.94), it was comparable in its discrimination to the total number of melancholic items (AUC=0.93, CI=0.03–0.01, $p<0.38$) and a difference score (AUC=0.93, CI=0.03–0.01), $p<0.44$). As the 'difference' score was quantified here by ROC analysis as 4 or more for putative melancholia – and thus had the same cut-off as quantified for the self-report measure – and had high discrimination (comparable to two other highly discriminating strategies), we viewed it as the optimal identified strategy, and therefore tested that strategy in our shortly reported validity analyses. The 'difference score' of 4 or more was quantified as having a sensitivity of 0.84, specificity of 0.92, positive predictive value of 0.90 and negative predictive value of 0.88.

3.5. Examining the validity of the derived SERDEX 'difference' measures

Ascriptions for melancholia (see Parker and Hadzi-Pavlovic, 1996a) as against non-melancholic depression include an older age and older age at onset, greater likelihood of a family history, decreased likelihood of disordered personality and life event stressors to onset, and greater responsivity to antidepressant drugs and ECT. We therefore examined (see Table 3) the extent to which our data set allowed examination of any such validating variables for those assigned as melancholic or non-melancholic by SERDEX 'difference' scores (i.e., a difference score of 4 or more being indicative of melancholia).

In relation to the self-report SERDEX measure, those who scored 4 or more (i.e., putative melancholia) did not differ by age, family history variables, a measure of developmental difficulties (i.e., difficulties with parents) or by rates of co-morbid drug or alcohol problems. They did score higher on two state depression measures (significantly for the DMI-10), reported a significantly longer current depressive episode, returned a higher global functioning score (indicating greater impairment), and reported fewer current and lifetime stressors. They were less likely to have one or more lifetime anxiety disorders, showing lower rates of social anxiety and obsessive compulsive disorder (OCD) in particular. On the MAP-incorporated Temperament and Personality measure (Parker and Manicavasagar, 2005) assessing personality styles viewed as predisposing to depression, they did not differ in returning lower scores on any personality style scale (actually scoring higher on 'personal reserve'), while they also failed to differ on the two scales assessing disordered personality functioning (i.e., low cooperativeness and effectiveness). They were, however, more likely to report having received electroconvulsant therapy (ECT).

In relation to the clinician-rated SERDEX measure, those who scored 4 or more (i.e., putative 'melancholia') had a longer duration for their current depressive episode but did not differ on state depression measures or on the global functioning score. They were significantly more likely to have received ECT. They were less likely to report a history of difficulty with parents, reported fewer current and lifetime stressors and were less likely to have a lifetime diagnosis of one or more anxiety disorders, with the strongest differences observed specifically with respect to social phobia, generalised anxiety disorder and OCD. On the 10 Temperament and Personality scales, so assigned melancholic patients were significantly more likely to return lower scores on five of eight scales (i.e., irritability, anxious worrying, self-criticism, interpersonal sensitivity and self-focused) and also rate significantly higher on effectiveness and cooperativeness, suggesting lower levels of personality dysfunction.

4. Discussion

Melancholia is variably positioned as a more 'severe' form of depression or as a separate 'type.' If the former position is valid

Table 3
Comparisons of those assigned by a cut-off score of 4 or more to melancholic (and non-melancholic) depression by the self-rated and clinician-rated SERDEX measure (difference score of 4 or more for melancholia) against validator variables.

MAP variable	Self-rated SERDEX $n=278$			Clinician-rated SERDEX $n=278$		
	Mel ($n=121$)	Non-mel ($n=157$)	Sig.	Mel ($n=114$)	Non-mel ($n=164$)	Sig.
Sex-female (%)	55.0	48.7	0.30	49.1	53.0	0.52
Age at assessment	41.9	40.0	0.23	42.1	40.4	0.29
Age at depression onset	21.8	20.5	0.37	22.3	20.5	0.24
Current depression (days)	76.7	43.5	<0.001	72.4	55.2	<.001
DMI-10 total score	22.1	19.0	<0.001	21.0	20.1	0.35
QIDS total score	16.0	14.9	0.33	15.2	15.6	0.33
Global functioning total score	15.4	12.7	<0.001	14.5	13.6	0.19
Family history of depression (%)	66.9	72.7	0.28	67.5	71.8	0.44
Family history of bipolar (%)	17.9	16.7	0.77	13.2	20.1	0.13
Family history of alcoholism (%)	37.5	36.7	0.90	34.2	39.0	0.41
Difficulty with parents (%)	38.3	48.7	0.084	36.8	49.4	0.038
Drug problem (%)	8.3	13.9	0.15	7.9	14.0	0.12
Alcohol problem (%)	15.0	19.6	0.32	13.2	20.7	0.10
Lifetime Anxiety Disorders (%)						
Social phobia	20.8	36.7	0.004	17.5	38.4	<0.001
GAD	27.5	33.5	0.28	23.7	36.0	<0.029
Panic	26.7	22.2	0.38	22.8	25.0	0.67
Agoraphobia	15.8	18.4	0.58	12.3	20.7	0.067
OCD	15.0	26.6	0.020	14.0	26.8	0.011
PTSD	14.7	17.1	0.60	11.7	19.1	0.11
1 or more lifetime anxiety disorders	51.7	64.6	0.030	46.5	67.7	<0.001
Total no. current stressors	2.2	3.2	<0.001	1.9	3.4	<0.001
Total no. lifetime stressors	4.3	5.4	<0.001	4.2	5.5	<0.001
Tricyclics—ever used (%)	20.8	21.5	0.89	25.4	18.3	0.15
ECT—ever received (%)	8.3	0.6	0.001	7.9	1.2	0.005
Personality Styles						
Social avoidance	12.2	11.2	0.11	11.4	11.9	0.46
Irritability	11.0	10.4	0.45	9.5	11.6	<0.05
Perfectionism	16.8	18.0	0.07	17.5	17.3	0.70
Anxious worrying	14.3	14.6	0.70	12.8	15.6	<0.001
Personal reserve	11.0	9.3	<0.05	9.2	10.7	0.06
Self-criticism	16.1	15.5	0.31	14.9	16.4	<0.05
Interpersonal sensitivity	8.6	9.1	0.43	7.2	10.1	<0.001
Self-focused	4.1	3.7	0.39	3.3	4.3	<0.005
Cooperativeness	23.0	23.2	0.67	24.4	22.3	<0.001
Effectiveness	14.9	15.9	0.18	16.6	14.4	<0.005

then it might be simply quantified on appropriate measures of severity and differentiated from non-melancholic clinical depression. If it is a distinct 'type', its formal clinical definition is important if there are aetiological and/or treatment implications. Assuming the latter model, then why has it proved resistant to clear definition? Two principal reasons for its ineffability have been offered and must be respected in developing measurement approaches. First, that it is more a 'system' (rather than focal) disorder involving disruptions in functional neurocircuits, particularly involving frontal-subcortical pathways and varying neurotransmitter contributions. The site(s) of such disruptions and the contribution of differing neurotransmitters might all contribute to differential symptoms across differing patients—as occurs for Parkinson's disease, and with the similarities and overlaps between melancholia and Parkinson's disease being considered in depth by Austin and Mitchell (1995). Second, and as detailed in the Introduction, its symptoms and their measurement are difficult to operationalise clearly, making reliance on symptom measurement alone problematic. Such a problem is not unique to melancholia. Parkinson's disease also lacks a gold standard laboratory diagnostic test, but this does not argue against neurologists seeking to define the condition and differentiate it from other salient conditions—assessing both clinical symptoms and illness correlates. Thus, the rationale for our development and evaluation of the previously named SERDEX measure—pursuing delineation of melancholia using multiple symptom and course of illness 'signals', while respecting their intrinsic fuzziness and that, at best, only prototypic (rather than pristine) definition might be achievable.

We chose candidate descriptors that reflected historical ascriptions and which we had refined from previous studies over the last two decades. One of the higher-order ascriptions to melancholia is that it is more 'biological' in its origins with certain over-represented rather than specific symptoms (e.g., anergia, anhedonia), while the residual non-melancholic disorders (once termed 'reactive' and 'neurotic' depression) are more generally positioned as reflecting the impact of stress with or without a contribution of a predisposing personality style. Such a model framed our development and validation strategies.

In this journal we previously reported (Parker et al., 2012) the utility of the SERDEX measure in a relatively small sample of 141 patients and as a self-report strategy only, but there demonstrated that context and course of illness variables were more differentiating than symptom variables. In this report the sample was doubled and we examined the properties of both self-report and clinician-rated versions. A key study limitation was the accuracy of assigning a diagnosis of melancholic versus non-melancholic depression—in that it weighted clinical judgment (and therefore had attendant limitations). A second limitation was the circularity of our logic. Our clinicians weighted certain symptoms and illness correlates in making a diagnosis of melancholia and which alone might have ensured their over-representation in those assigned a diagnosis of 'melancholia' and under-representation in the residual non-melancholic class. Such

limitations can only be conceded in the absence of independent raters and a benchmark diagnostic standard.

For both the self-report and clinician-rated SERDEX items forms, all 24 items had higher prevalences in their putative diagnostic (i.e., clinically defined melancholic or non-melancholic) group, but with not all differences formally significant. Chi square and odds ratio analyses identified greater specificity and more significant differences in the clinician-rated than the self-report SERDEX item set, a finding that could reflect the circularity issue noted in the previous paragraph or clinicians being more valid raters. Agreement between self-report and clinician ratings was weak, indicating that one strategy was likely to be superior to the other. The 'spread' of self-report SERDEX responses by non-melancholic patients across the five A–B prototypic patterns was also noteworthy and may reflect diagnostic imprecision or a rater bias. In relation to the last, we suspect that some patients may have interpreted the form as indicating that the Descriptor A set captured a more 'serious' depression and so preferentially rated items that would affirm the severity of their perceived condition to the clinician for assessment. If a valid interpretation, it adds to arguments favouring the clinician-rated version above the self-report version. Conversely, the 'success' of the clinician-rated measure may also reflect a rater bias, as noted in the previous paragraph.

Irrespective of the rater strategy, we demonstrated that the most discriminating items were equally likely to be symptoms or course of illness variables, and thus arguing against reliance on symptoms only. Similarly, the five-point prototype measure (asking whether the depressive picture corresponded more closely to the Item A or Item B set) showed greater differentiation across the clinician-rated rather than the self-rated version. In relation to the former, specificity was impressive in relation to the clinically diagnosed melancholic patients (with nearly 90% rated as matching or being closer to Description A and only 3% to Description B). Specificity across both self-report and clinician-rated approaches was less impressive for those receiving a clinical diagnosis of a non-melancholic depression, with matching or closeness to Description B being achieved by 67% of those non-melancholic patients assigned by clinician ratings and 29% of those using the self-report measure.

A key set of ROC analyses quantified whether the self-rated SERDEX measure approach (i.e., symptoms plus illness correlates) was superior to self-rating of symptoms only (as assessed by our comprehensive 32-item SDS measure). All four self-report SERDEX applications were superior to use of the SDS (symptoms only) approach, and this again underlies what is perhaps the most important study finding.

We formally quantified that an item 'difference' score (i.e., subtracting the affirmed number of Descriptor B items from the affirmed number of Descriptor A items) was the most discriminating—having a sensitivity of 0.69 and specificity of 0.77. A set of ROC analyses indicated that, for the clinician-rated SERDEX approaches, both the 'prototype' and 'difference' scores effected the greatest discrimination (i.e., respective sensitivities of 0.90 and 0.84, and specificities of 0.90 and 0.92). While such analyses generate optimised cut-off scores (respecting nuances of the sample under analysis) and would be unlikely to be quantified at such high levels in replication studies, the key conclusion here is again in indicating that the clinician-rated version of the SERDEX measure appeared distinctly superior to the self-rated version. As the 'difference' score was the most discriminating in analyses of the self-report SERDEX and close to being the most discriminating in the clinician-rated SERDEX, we favour its use.

Finally, we sought to examine the validity of the SERDEX measures by directly comparing the extent to which optimal 'difference' scores identified 'melancholic' and 'non-melancholic' groups whose constituents met the ascriptions accorded to those depressive sub-types. While validation was limited (in that we did not demonstrate anticipated distinct differences in family history, age and depression severity), once again the clinician-rated SERDEX version generated greater empirical support than did the self-rated version. So assigned non-melancholic patients had distinctly higher rates of anxiety, current and lifetime stressors, early parental difficulties, at-risk personality styles and disordered personality functioning, consistent with the attribution that non-melancholic depression is commonly underpinned by stress and personality causal factors.

It is important to note that assessment of symptoms by the measure focuses on their presence during a previous or current (as here) episode and not on assessing their severity. This is of advantage if melancholic and non-melancholic depressive conditions are to be distinguished by their differential clinical pattern and not by severity per se. Our studies have been undertaken in clinical samples only to this stage, and it would be important for future replication studies to be undertaken in community samples. While the clinician-rated version of the measure was superior in this study, there would be wisdom in examining both the self-report and clinician-rated versions in any such community studies although we would argue for any assessment of the latter version involving raters with clinical skills.

As the SERDEX acronym incorporates its original emphasis on a self-report measure and the data distinctly favour the clinician-rated measure, the measure requires renaming. Respecting some history and its correspondence to the model employed in the Newcastle Index (Carney et al., 1965) we have elected to re-name it the Sydney Melancholia Prototype Index (SMPI), with comparable self-report and clinician-rated versions. We suggest that the measure is worthy of further evaluation in definitional, causal and treatment studies. Its salience may best lie in that its underlying model approximates to a clinician's approach to formulating a depressive 'pattern' in evaluating symptoms, illness correlates and potential risk factors. As noted by McHugh and Slavney (2012), psychiatry might benefit from addressing psychiatric disorders "in the same way that internists address physical disorders, explaining the clinical manifestations as products of nature to be comprehended not simply by their outward show but by the causal processes and generative mechanisms known to provoke them".

Role of funding source

This study was supported by an NHMRC program grant (510135). The NHMRC did not participate in the study design, nor in the data collection and analysis process, or in any phase of the manuscript preparation.

Conflict of interest

Nothing to declare.

Acknowledgments

This study was funded by an NHMRC Program Grant (510135) and an Infrastructure Grant from the NSW Department of Health.

References

Austin, M.P., Mitchell, P., 1995. The anatomy of melancholia: does frontal-subcortical pathophysiology underpin its psychomotor and cognitive manifestations? Psychological Medicine 25, 665–672.

Berrios, G.E., 1988. Melancholia and depression during the 19th century: a conceptual history. British Journal of Psychiatry 153, 298–304.

Carney, M.W.P., Roth, M., Garside, E.R.F., 1965. The diagnosis of depressive syndromes and the prediction of E.C.T. response. British Journal of Psychiatry 111, 659–674.

Jackson, S.W., 1986. Melancholia and Depression: From Hippocratic Times to Modern Times. Yale University Press, New Haven, CT.

McHugh, P.R., Slavney, P.R., 2012. Mental illness—comprehensive evaluation ofr checklist? New England Journal of Medicine 366, 1853–1855.

Parker, G., Fletcher, K., Barrett, M., Synnott, H., Breakspear, M., Rees, A.M., et al., 2010. Inching toward Bethlehem: mapping melancholia. Journal of Affective Disorders 123, 291–298.

Parker, G., Fletcher, K., Hadzi-Pavlovic, D., 2012. Is context everything to the definition of clinical depression? A test of the Horwitz and Wakefield postulate. Journal of Affective Disorders 136, 1034–1038.

Parker, G., Fletcher, K., Hyett, M., 2008. The mood assessment program: a computerised diagnostic tool for deriving management plans for mood disorders. Medical Journal of Australia 188, S126–S128.

Parker, G., Hadzi-Pavlovic, D., 1996a. Melancholia: A Disorder of Movement and Mood. Cambridge University Press, NY.

Parker, G., Hadzi-Pavlovic, D., 1996b. Development and structure of the CORE system. In: Parker, G., Hadzi-Pavlovic, D. (Eds.), Melancholia: A Disorder of Movement and Mood. Cambridge University Press, NY, pp. 82–129.

Parker, G., Hadzi-Pavlovic, D., Boyce, P., 1996. Issues in classification: II. Classifying melancholia. In: Parker, G., Hadzi-Pavlovic, D. (Eds.), Melancholia: A Disorder of Movement and Mood. Cambridge University Press, NY, pp. 20–37.

Parker, G., Hilton, T., Bains, J., Hadzi-Pavlovic, D., 2002. Cognitive-based measures screening for depression in the medically ill: the DMI-10 and the DMI-18. Acta Psychiatr Scand journal 105, 419–426.

Parker, G., Manicavasagar, V., 2005. Modelling and Managing the Depressive Disorders: A Clinical Guide. Cambridge University Press, NY.

Parker, G., Manicavasagar, V., Crawford, J., Tully, L., Gladstone, G., 2006. Assessing personality traits associated with depression: the utility of a tiered model. Psychological Medicine 36, 1131–1139.

Rush, A.J., Trivedi, M.H., Ibrahim, H.M., Carmody, T.J., Arnow, B., Klein, D.N., et al., 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.

Taylor, M.A., Fink, M., 2006. Melancholia: The Diagnosis, Pathophysiology, and Treatment of Depressive Illness. Cambridge University Press, Cambridge.

Widlöcher, D.J., 1983. Psychomotor retardation: Clinical, theoretical, and psychometric aspects. Psychiatric Clinics of North America 6, 27–40.