The DSM-IV definition of severity of major depression: inter-relationship and validity

V. Lux[1], S. H. Aggen[2,3] and K. S. Kendler[2,3,4]*

[1] *Department of Psychology, Free University Berlin, Germany*
[2] *Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA*
[3] *Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA*
[4] *Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA*

Background. Severity is an important characteristic of major depression (MD) and an 'episode specifier' in DSM-IV classifying depressive episodes as 'mild', 'moderate' or 'severe'. These severity subtypes rely on three different measures of severity: number of criteria symptoms, severity of the symptoms and degree of functional disability. No prior empirical study has evaluated the coherence and validity of the DSM-IV definition of severity of MD.

Method. In a sample of 1015 (518 males, 497 females) Caucasian twins from a population-based registry who met criteria for MD in the year prior to interview, factor analysis and logistic regression were conducted to examine the inter-relationships of the three severity measures and their associations with a wide range of potential validators including demographic factors, risk for future episodes, risk of MD in the co-twin, characteristics of the depressive episode, the pattern of co-morbidity, and personality traits.

Results. Correlations between the three severity measures were significant but moderate. Factor analysis indicated the existence of a general severity factor, but the factor was not highly coherent. The three severity measures showed differential predictive ability for most of the validators.

Conclusions. Severity of MD as defined by the DSM-IV is a multifaceted and heterogeneous construct. The three proposed severity measures reflect partly overlapping but partly independent domains with differential validity as assessed by a wide range of clinical characteristics. Clinicians should probably use a combination of severity measures as proposed in DSM-IV rather than privileging one.

Introduction

Severity is an important characteristic of major depression (MD), predicting short-term treatment outcomes (Blom *et al.* 2007), probability of recovery (Rubenstein *et al.* 2007), response to pharmacological treatment (Angst *et al.* 1995; Kasper *et al.* 1997; Hirschfeld, 1999), probability of suicidal ideation (Alexopoulos *et al.* 1999) and length of depressive episode (Kennedy *et al.* 2004; Melartin *et al.* 2004). In the DSM-IV criteria for MD (APA, 1994), severity is the first of the 'episode specifiers' providing the clinician with the ability to classify episodes as 'mild', 'moderate' or 'severe'. To our knowledge, the definition of severity in DSM-IV ('Severity is judged to be

mild, moderate or severe based on the number of criteria symptoms, the severity of the symptoms and the degree of functional disability and distress') derives from expert opinion and was neither empirically developed nor subsequently validated.

The aim of this report is to contribute to an empirical validation of the DSM-IV definition of severity of MD by evaluating its coherence and investigating its inter-correlations and associations with clinically relevant phenomenon. To do so, we examine individuals who met criteria for MD in the past year from the large epidemiological sample of the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD; Kendler & Prescott, 2006). We first examine the inter-correlations of three severity measures used in the DSM-IV: 'number of criteria symptoms', 'the severity of the symptoms' and 'degree of functional disability'. Next, we test the relationships of these measures with a set of wide-ranging potential validators including demographic factors, risk for

* Address for correspondence: Dr K. S. Kendler, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Box 980126, Richmond, VA 23298-0126, USA.

(Email: kendler@vcu.edu)

future episodes, the risk of MD in the co-twin, characteristics of the depressive episode, co-morbidities, and personality traits.

Method

Sample

Participants in this report derive from two inter-related studies in Caucasian same-sex twin pairs who participated in the VATSPSUD (Kendler & Prescott, 2006). All subjects for the VATSPSUD were ascertained from the Virginia Twin Registry, a population-based register formed from a systematic review of birth certificates in the Commonwealth of Virginia. Female–female (FF) twin pairs, from birth years 1934–1974, became eligible if both members previously responded to a mailed questionnaire in 1987–1988, to which the response rate was about 64%. The first face-to-face interview (FF1) was completed by 92% ($n = 2163$) of the eligible twins. These twins participated in three subsequent interviews with cooperation rates ranging from 85% to 93%. Data on the male–male and male–female (MMMF) pairs came from a sample (birth years 1940–1974) initially ascertained directly from registry records, which contained all twin births, by a telephone interview to which the response rate was 72% ($n = 6812$). This sample was re-interviewed once with an 83% response rate. Zygosity was determined by discriminate function analyses using standard twin questions validated against DNA genotyping in 269 FF and 227 MM pairs (Kendler & Prescott, 1999).

Assessment of depression

MD diagnoses and the corresponding severity measures were based on the 'last year prevalence' module in the FF1 and MMMF1 interviews. In this section, every subject was asked individually whether they experienced each of the disaggregated criteria symptoms for DSM-IV MD in the year prior to the interview. By disaggregated, we mean that they were asked separate questions for psychomotor agitation or retardation, insomnia or hypersomnia, weight loss or gain, and appetite increase or decrease.

The DSM-IV criteria for MD were met by 217 twins from the FF1 and 798 twins from the MMMF1 interviews. Of these 1015 twins, 518 were males and 497 were females. At the time of the first interview, their age ranged from 18 to 57 years with a mean of 34.5 years. There were 83 twin pairs with both twins diagnosed with MD. In addition, two more pairs from two triplets were diagnosed with MD.

We assessed the severity of each endorsed symptom, using three approaches for different symptom groups. For those symptoms with a 'natural metric' (e.g. hours of sleep, pounds of weight), we asked the subject how much that had changed. For appetite change and psychomotor agitation, the interviewer asked directly how severe was the 'appetite decrease' and the 'restlessness', recording the twins' response on a three-point scale ('severe', 'moderate' and 'mild'). For all other symptoms that had no such natural metric (e.g. feelings of sadness, loss of concentration, worthlessness/guilt), the interviewer asked how much the symptom interfered with daily life activities. Responses were recorded on a four-point scale ('completely', 'a lot', 'some' or 'hardly at all'). We also asked a question about the etiology of the symptom that permitted us to exclude those due to medication or illness. In addition, the respondents answered three questions about how much their work (or housework if homemaker), leisure time activities and interpersonal relationships were interfered with or impaired by the worst depressive episode experienced in the year prior to the interview. Responses were on a three-point scale ('severe', 'moderate' and 'none').

Severity indices

We operationalized the DSM-IV 'number of criteria symptoms' (hereafter 'criteria count') as the number of DSM-IV 'A criteria' met by these individuals. This ranged from 5 to 9. For the disaggregated symptoms, if an individual met at least one (e.g. weight loss), then the entire criteria ('appetite/weight change') were counted as positive.

We operationalized the DSM-IV 'severity of the symptoms' (hereafter 'symptom severity') using the severity measures outlined above. For the disaggregated criteria (e.g. weight, sleep and psychomotor changes), we used a 'most severe' rule. To compare the ordinal scores with actual pounds for the weight items or hours for the sleep items, we transformed the metric measures for weight/hours into an ordinal scale, correcting the weight change for total reported body weight. If a symptom was not reported or a symptom was reported as due to medication or illness, the impairment question was coded as 'missing'. The factor analyses were run in Mplus (Muthen & Muthen, 2004), which allowed the use of all available observations despite missing values, given that no severity measure was present if a symptom was not endorsed.

To create a measure comparable to symptom severity, we operationalized the DSM-IV 'degree of functional disability' (hereafter 'syndromal impairment') as the factor score derived from our three questions measuring occupational, social and

relational impairment resulting from the depressive episode. In the initial description of this severity specifier, DSM-IV writes 'degree of functional disability and distress'. However, in the subsequent text providing the specifics of the mild, moderate and severe subtypes, distress goes unmentioned. Therefore, our main analyses focused solely on our impairment measures. Further discussion of this issue is presented in the limitations section below.

Validators

The VATSPSUD includes a rich set of data about future episodes, depressive episodes of the co-twin, lifetime co-morbidities, demographic characteristics, and characteristics of the index depressive episode (Kendler & Prescott, 2006). For demographic characteristics, characteristics of the index depressive episode and last year co-morbidity with general anxiety disorder (GAD), the data came from the same interview wave. For depressive episodes of the co-twin and all other co-morbidities, data were obtained from all interview waves to include the best amount of accessible information in the analysis. To reduce potential confounding effects of unequal follow-ups, future depressive episodes were diagnosed based on the 'last year prevalence module' of the second interview.

GAD was diagnosed using the DSM-III-R criteria (APA, 1987), requiring a minimum of 1 month of duration. Lifetime panic disorder was also diagnosed using the DSM-III-R criteria. 'Any phobia' was diagnosed using an adaptation of DSM-III criteria (APA, 1980) requiring one or more unreasonable fears, including fears of different animals, social phobia and agoraphobia, that objectively interfered with the respondent's life. Nicotine dependence was defined as a score $\geqslant 7$ on the Fagerström Tolerance Questionnaire (FTQ; Fagerström, 1978), and alcohol dependence and illicit drug dependence were diagnosed using DSM-IV criteria (APA, 1994). Adult antisocial personality traits were defined as meeting $\geqslant 3$ of the DSM-III-R (APA, 1987) 'C criteria' for antisocial personality disorder. Extraversion was assessed with eight and neuroticism with 12 items from the short form of the self-administered Eysenck Personality Questionnaire (Eysenck *et al.* 1985). For 'co-occurring anxiety symptoms' we used a binary variable indicating whether the respondent endorsed at least one of two anxiety symptoms during the last 12 months in which they had their index depressive episode. These items were: 'felt anxious, nervous or worried' and 'muscles felt tense or felt jumpy or shaky inside'. 'Chronic MD' was defined as a depressive episode lasting $\geqslant 12$ months. For 'experiencing the MD out of the blue', we asked the respondent about their index episode

whether 'something happened to make you feel that way or did the feeling just come on you "out of the blue"?' 'Seeking help' was assessed by a question asking whether the respondent went to get help from health professionals, ministers, self-help groups, or anyone else.

Data analysis

We began by creating comparable measures of our three severity indices. For symptom severity, where we had nine variables, we used a confirmatory factor analysis (CFA) carried out in Mplus accounting for the non-independence of the twin data and using a weighted least square estimation method based on polychoric correlations (Muthen & Muthen, 2004). We also used the pairwise deletion function in Mplus, which allows the inclusion of all observations in the factor analysis, even if there are missing data for some of the items if the symptoms were not endorsed as present. Fit was assessed by two indices, the Comparative Fit Index (CFI) and the Tucker–Lewis Index (TLI) (Bentler, 1990), where values >0.95 indicate a good fit to the data.

Our measure of syndromal impairment and our overall measures of severity each contained only three variables so a CFA was not feasible. Instead, we carried out, for both these analyses, an exploratory factor analysis (EFA) in SAS (SAS Institute, 2005) using a polychoric correlation matrix and an unweighted least square estimation method. No rotation was possible so the loadings on the single factor are presented.

We evaluated the performance of our three severity indices by their relationship to a range of potential validators. Depending on the distributional properties of the validator variable, these analyses were conducted using binary or cumulative logit models in the LOGISTIC function in SAS (SAS Institute, 2005). The severity index was the predictor and the validator variable the dependent variable, with age and sex included as covariates. For the validator 'MD diagnosis in co-twin', zygosity was added as a covariate in the model.

We then explored the unique predictive power of each of our severity indices using the GENMOD procedure in SAS (SAS Institute, 2005). Our approach involved examining pairs of our severity indices in logistic regression. If both indices were significantly associated with the validator, we would start with severity index1 (and age and sex) and then added severity index2 to the model. If the fit of the model significantly improved, then index2, for this validator, explained additional variance not captured by index1. To confirm this finding, we then repeated the analyses the other way around; that is, showing that the

Table 1. *Factor loadings in confirmatory factor analysis (CFA) of symptom severity: the one-factor solution*

Item	Factor loadings
Loss of interest	0.81
Depressed mood	0.78
Feeling worthless/guilty	0.72
Loss of concentration	0.72
Fatigue	0.65
Thoughts of death/suicidal ideation	0.62
Psychomotor agitation/retardation	0.47
Insomnia/hypersomnia	0.36
Appetite/weight changes	0.35

addition of index1 to a model with index2 significantly explained additional variance for the validators. If, however, only one of the two indices was statistically significant, then we only required the addition of the significant to the non-significant index in the model to show a significant improvement in fit. Finally, if none of the indices was statistically significant, we started with the index with the lower odds ratio (OR) and added the index with the higher OR to test if the improvement was statistically significant. p values are reported two-tailed except for risk of MD in co-twin, where we report one-tailed values, given the prior prediction of twin resemblance.

Results

Factor analysis of symptom severity

We fitted, using a CFA, one- and two-factor oblique solutions. The one-factor solution produced a good fit [CFI = 0.96, TLI = 0.97, root mean square error of approximation (RMSEA) = 0.07]. Although a two-factor solution also explained the data well (CFI = 0.97, TLI = 0.97, RMSEA = 0.06), the resulting factors were too highly correlated (+0.83) to be meaningfully separable. Therefore, we used the one-factor solution (Table 1). The highest loadings were seen for the three 'cognitive' criteria of loss of interest, sad mood and feelings of worthlessness. All criteria loaded in excess of +0.40 with the exception of sleep and appetite/weight changes.

Factor analysis of syndromal impairment

A factor analysis of these three features of syndromal impairment ($n = 1005$) produced a single coherent factor with the following loadings: impairment in leisure time activity +0.79, impairment in relationships +0.60 and occupational impairment +0.57.

The three measures of severity of depression: inter-correlation and factor analysis

Although highly significant, Pearson product-moment correlations between the three severity indices were modest: criteria count and syndromal impairment +0.25 ($n = 1005$, $p < 0.0001$), criteria count and symptom severity +0.37 ($n = 1015$, $p < 0.0001$), and syndromal impairment and symptom severity +0.40 ($n = 1005$, $p < 0.0001$). Factor analysis produced a single 'severity' factor, with moderate loadings: symptom severity (+0.75), syndromal impairment (+0.52) and criteria count (+0.51).

Logistic regression analysis of severity indices

Table 2 shows the association between these three severity indices of MD [criteria count (CC), syndromal impairment (SI) and symptom severity (SS)] and 23 wide-ranging potential validators available in the VATSPSUD. ORs, p values and 95% confidence intervals (CIs) are presented controlling only for age and sex. A p value <0.05 was considered significant, indicating that the finding was not likely to be a chance effect. For the cumulative logit models, the ORs of a one standard deviation (S.D.) increase of the dependent variable are presented in the table, and also a parallel result for the general severity factor.

Of the many results presented here, seven are noteworthy. First, at a global level, criteria count and symptom severity were each significantly associated with 14 validators and syndromal impairment with 12. The mean (S.D.) ORs for all these three indices were: criteria count 1.27 (0.24), syndromal impairment 1.31 (0.27) and symptom severity 1.31 (0.25). Second, syndromal impairment was most strongly associated with lifetime co-morbidities with anxiety disorders, symptom severity with substance use disorders and criteria count with antisocial personality traits. Third, regarding our two personality measures, high levels of neuroticism were most strongly associated with symptom severity, whereas syndromal impairment was the index most strongly associated with low extraversion. Fourth, regarding features of the current episode, criteria count was the index most strongly associated with prominent concurrent anxiety and symptom severity was most strongly associated with duration and help-seeking, whereas syndromal impairment was most strongly associated with a chronic episode and the occurrence of the MD 'out of the blue'.

Fifth, none of the severity criteria were significantly associated with the two measures obtained of lifetime MD, age at first onset and number of lifetime episodes. Sixth, with respect to demographic features, syndromal impairment was most strongly related to younger age at current episode, whereas symptom severity was most robustly related to sex (more severe in males). None of the severity measures were significantly associated with being married/living with partner, low family income or years of education. Seventh, symptom severity most strongly predicted future depressive episodes, whereas only criteria count was significantly associated with risk of MD in the co-twin.

The last three columns of Table 2 summarize the results of the differential ability of our three indices of depressive severity to explain the variance of the validators; that is, if one measure of severity is in the model, does the inclusion of a second explain statistically significant additional variance for the specific validator? Of the 23 validators, criteria count and symptom severity explained statistically significant unique proportions of variance for 13, criteria count and symptom severity for 9, and syndromal impairment and symptom severity for 10.

Finally, the correlations in our measures of severity in the 83 twin and two triplet pairs in our sample concordant for a history of MD in the past year were: criteria count $+0.04$ ($p=0.35$), syndromal impairment $+0.09$ ($p=0.21$) and symptom severity $+0.20$ ($p=0.03$). The general severity factor was also modestly correlated in these pairs ($+0.22$, $p=0.02$).

Discussion

The aim of this report was to evaluate empirically, for the first time to our knowledge, the DSM-IV definition of severity of MD. Our analysis shows that this construct was neither simple in structure nor uniform in validity. Four specific findings are noteworthy. First, the correlations between the three DSM-IV indices of depressive severity were only moderate in magnitude. Taking into account that symptom severity and overall syndromal impairment partly overlap in content, this finding is even more striking. In addition, when examined together, the three severity indices did not form a highly coherent factor. Second, the individual measures of severity and also the general severity factor were validated in the sense that their association to a fairly wide range of characteristics in depressed patients was examined, with none of these validators playing any role in the diagnostic process. Classifying depressed subjects by severity can tell you some important things about the expected patterns of co-morbidity, other clinical features and prognosis.

Third, the patterns of relationships between the severity indices and our set of validators differed meaningfully across the three indices. Fourth, in most of the cases (17 out of 23), at least one severity index explained significantly distinct proportions of variance of our validators when added to a model with one of the other indices. That is, these three different measures of depressive severity were often associated with different things. In summary, these results suggest that, as operationalized in DSM-IV, the concept of severity of MD is best understood as a multifaceted heterogeneous construct.

Our findings echo a principle articulated about schizophrenia more than 30 years ago by Strauss & Carpenter (1978): that symptoms and functional impairment in psychiatric illness are only loosely interconnected. More recently, several studies focusing on MD have also reported only moderate correlations for various measures of impairment and criteria or symptom count (Kitamura *et al.* 1993; Faravelli *et al.* 1996; Huang *et al.* 2006). When higher correlations of depressive severity and impairment measures were reported, the authors either used a combination of criteria count and symptom severity to calculate the intercorrelations (Kroenke *et al.* 2001; Hiroe *et al.* 2005; Zimmerman *et al.* 2006) or compared syndromal impairment to overall severity of MD (Iannuzzo *et al.* 2006).

We were surprised at the low loadings of some of our measures of symptom severity on the common factor (e.g. appetite/weight and sleep). However, this has been seen in one other study (Olsen *et al.* 2003) and there was very limited evidence in our sample for a second distinct symptom severity factor. In addition, although not entirely comparable to our study, a weak performance of various disaggregated weight and sleep items as severity measures was also found in studies on different severity measures (Faravelli *et al.* 1996; Santor & Coyne, 2001; Zimmerman *et al.* 2006).

Specific findings in our sample for inter-relationships between the three indices of depressive severity and a range of external validators also has precedent in the literature. Prior studies have reported, for example, that impairment is related to risk for future depressive episodes (Rodriguez *et al.* 2005), comorbidities with anxiety or substance use disorders (Mojtabai, 2001) and co-morbid panic-depression (Roy-Byrne *et al.* 2000); and that impairment is not related to sex (Sheehan *et al.* 1996) or age of onset of depression (Zisook *et al.* 2004). In addition, our finding that all three severity indices were significantly associated with chronic depression also corresponds to earlier findings (e.g. Pettit *et al.* 2009). In contrast to our results of males reporting higher symptom

Table 2. *A comparison of the three DSM-IV severity indices for major depression on a range of potential validators*

Dependent variable	Logit model	n (n')	Criteria count (CC) OR (95% CI)	Syndromal impairment (SI) OR (95% CI)	Symptom severity (SS) OR (95% CI)	General severity factor OR (95% CI)	CC – SI	CC – SS	SI – SS
Co-morbidities									
Generalized anxiety disorder (past year)	BL	1015 (1005)	1.13 (0.98–1.31)	1.28* (1.02–1.60)	1.23 (0.99–1.54)	1.33* (1.06–1.68)	N.S.	N.S.	N.S.
Panic disorder	BL	859 (850)	1.44** (1.14–1.80)	1.85** (1.27–2.68)	1.60** (1.14–2.22)	1.85** (1.31–2.62)	*	N.S.	N.S.
Any phobia	BL	881 (872)	1.28*** (1.15–1.43)	1.55*** (1.31–1.84)	1.31** (1.12–1.53)	1.56*** (1.32–1.85)	**	N.S.	N.S.
Alcohol dependence	BL	880 (871)	1.23** (1.10–1.38)	1.02 (0.86–1.21)	1.29** (1.09–1.52)	1.32** (1.11–1.58)	††	N.S.	††
Nicotine dependence	BL	558 (551)	1.31** (1.13–1.51)	1.17 (0.95–1.45)	1.40** (1.13–1.72)	1.52** (1.22–1.89)	††	N.S.	††
Any illicit drug dependence	BL	844 (835)	1.23** (1.06–1.44)	1.21 (0.94–1.54)	1.33* (1.07–1.65)	1.41** (1.12–1.78)	†	N.S.	†
Adult antisocial personality traits	BL	752 (745)	1.35*** (1.17–1.57)	1.25 (0.99–1.57)	1.26* (1.02–1.54)	1.47** (1.17–1.83)	††	N.S.	N.S.
Personality traits									
Higher neuroticism score	CL	977 (967)	1.75*** (1.55–1.98)	1.80*** (1.49–2.18)	1.81*** (1.51–2.16)	2.33*** (1.92–2.83)	***	**	***
Lower extraversion score	CL	977 (967)	1.06 (0.97–1.16)	1.26** (1.09–1.44)	1.08 (0.95–1.23)	1.16* (1.01–1.34)	††	N.S.	††
Characteristics of index depressive episode									
Prominent anxiety symptoms co-occurring	BL	992 (982)	1.79*** (1.48–2.16)	1.62*** (1.27–2.05)	1.45** (1.16–1.81)	1.97*** (1.52–2.55)	**	N.S.	*
Duration depressive episode in weeks	CL	1015 (1005)	1.68*** (1.48–1.92)	1.50*** (1.23–1.84)	1.70*** (1.41–2.04)	2.10*** (1.72–2.57)	*	**	N.S.
Chronic MD	BL	1015 (1005)	1.63*** (1.42–1.87)	1.71*** (1.37–2.12)	1.70*** (1.40–2.06)	2.15*** (1.74–2.67)	**	**	**
'Out of the blue'	BL	1015 (1005)	1.27** (1.09–1.46)	1.34* (1.06–1.69)	1.05 (0.85–1.30)	1.25 (0.99–1.56)	N.S.	††	†
Seeking help	BL	1014 (1004)	1.32*** (1.19–1.47)	1.61*** (1.37–1.89)	1.66** (1.42–1.94)*	1.92*** (1.62–2.27)	***	**	**
Prior depression history									
(Younger) Age of first onset of MD	CL	719 (712)	1.03 (0.91–1.17)	1.02 (0.85–1.23)	0.95 (0.79–1.13)	0.99 (0.82–1.20)	N.S.	N.S.	N.S.
Number of episodes before index episode	CL	715 (708)	1.04 (0.93–1.15)	1.05 (0.90–1.24)	1.04 (0.90–1.21)	1.07 (0.91–1.26)	N.S.	N.S.	N.S.
Demographic characteristics at index episode									
Age (being younger)	CL	1015 (1005)	0.98 (0.90–1.08)	1.24** (1.08–1.43)	1.20** (1.05–1.36)	1.22** (1.06–1.41)	††	††	N.S.
Sex (being male)	BL	1015 (1005)	1.02 (0.92–1.13)	0.94 (0.80–1.09)	1.22** (1.06–1.41)	1.16 (0.99–1.35)	N.S.	††	*

		n (n')						
Not married or living with partner	BL	1015 (1005)	1.05 (0.95–1.16)	0.95 (0.81–1.11)	0.98 (0.85–1.14)	0.99 (0.84–1.16)	N.S.	N.S.	N.S.
Less family income	CL	1015 (1005)	1.09 (1.00–1.20)	1.05 (0.91–1.21)	1.06 (0.93–1.21)	1.09 (0.95–1.26)	N.S.	N.S.	N.S.
Less school years	CL	1015 (1005)	0.93 (0.85–1.02)	0.96 (0.83–1.11)	1.10 (0.96–1.25)	1.10 (0.94–1.26)	N.S.	N.S.	N.S.
Other validators									
New episode of MD in second wave	BL	876 (867)	1.40** (1.25–1.57)	1.32** (1.11–1.57)	1.63*** (1.38–1.93)	1.78*** (1.49–2.13)	N.S.	**	N.S.
MD diagnosis in co-twin[a]	BL	801 (792)	1.12* (1.00–1.25)	1.10 (0.92–1.30)	1.02 (0.87–1.19)	1.10 (0.92–1.30)	N.S.	†	N.S.

MD, Major depression; BL, binary logit; CL, cumulative logit; covariates: age, sex; [a] one-tailed p values, additional covariate: zygosity. n, sample size; n', sample size of syndromal impairment and general severity factor (differences due to missing data).

For the CL models, the ORs for a one standard deviation (S.D.) increase in the dependent variable are presented.

*** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$: two-tailed p values/significant in both directions; ††† $p < 0.0001$, †† $p < 0.01$, † $p < 0.05$: significant in one direction; N.S., not significant.

---

severity, Scheibe *et al.* (2003) found no sex differences in severity of depression for interview-based measures. Our findings are also consistent with an earlier study on the same sample that found, using structural equation twin modeling, that the factors that impact on functional impairment in MD are partly separable from those that alter risk for the disorder (based on meeting sufficient DSM-IV criteria) (Foley *et al.* 2003).

The classification of the severity subtypes of MD in the ICD-10 clinical (WHO, 1992) and research criteria (WHO, 1993) differ in several ways from that proposed in DSM-IV: (i) the additional criterion 'loss of confidence and self-esteem', (ii) the use of 'type' of symptoms, especially somatic symptoms, as additional severity measures, and (iii) the inclusion of distress in the syndromal impairment in the clinical criteria. Despite these differences, our results carry at least two implications for the ICD-10 classification of a mild, moderate and severe depressive episode. First, by specifying, in both the clinical and research criteria, a minimum of symptoms for each severity subtype, the ICD-10 definition emphasizes criteria count as crucial to the overall assessment of severity of MD, an approach not entirely supported by our results. Second, surprisingly, syndromal impairment is included as part of the definition of depressive severity in the clinical (WHO, 1992) and not in the research criteria (WHO, 1993). This is not consistent with our own findings, where syndromal impairment explained unique proportions of variance as an index of depressive severity independent of symptom severity or criteria count.

There are several well-established depression scales providing valuable severity measures [e.g. the Hamilton Rating Scale for Depression (HAMD; Hamilton, 1960, 1967), the Beck Depression Inventory (BDI; Beck *et al.* 1961, 1996), the Montgomery–Äsberg Depression Rating Scale (MADRS; Montgomery & Äsberg, 1979), or the Zung Self-Rating Depression Scale (SDS; Zung, 1965)] that combine a symptom count and symptom frequency or intensity to form a sum score. The HAMD and the BDI also include a work impairment question. Validation studies suggest that the MADRS and the BDI are superior to the HAMD, especially the long version, as an index of depressive syndrome severity (e.g. Gibbons *et al.* 1993; Licht *et al.* 2005; Carmody *et al.* 2006). However, none of these measurements rely strictly on the DSM-IV definition of severity of MD. Either they are not restricted to the nine criteria A symptoms or they consider impairment and symptom severity as interchangeable and not parallel measures. Our data set did not contain any of these scales so we were unable to evaluate their performance. Of note, the notion

of unidimensionality of severity that these scales typically assume (see Gibbons *et al.* 1993) was not entirely supported by our results.

Limitations

These results should be interpreted in the light of five potentially important methodological concerns. First, our sample is limited to white twins born in the Commonwealth of Virginia and these results may or may not extrapolate to other samples. Second, the clinical characteristics we used as validators probably vary in the degree to which they reflect underlying severity, and so including some and excluding others could influence the general performance of the three severity measures. That is, the results of this comparison are necessarily limited to this particular set of validators.

Third, the nature of our analyses made it difficult to account formally, in most cases, for the non-independence of observations in our twin data. However, only about 17% of our data come from twin pairs, and correlations in all three severity measures in these pairs were fairly low ($\leqslant 0.20$). Thus, it is very unlikely that the twin character of our data influenced our results substantially. In addition, we explored formal corrections for the binary logit models and found no substantial effects. Fourth, our results could be affected by missing data regarding symptom severity. As the degree of symptom severity was obtained only when the symptom was endorsed, the problem of missing data reflects the inherent non-independence of symptom count and symptom severity and is unavoidable in this or any other similar analysis.

Fifth, as noted above, the DSM-IV is ambiguous about whether distress should be included in measures of the severity of MD. Although distress is included in the overall definition of severity as part of syndromal impairment, it is not further mentioned in the specification of the subgroups 'mild', 'moderate' and 'severe'. Therefore, our main analyses did not include distress ratings in our severity measures. To address whether our findings would change were we to incorporate measures of distress, we repeated in our MMMF subsample (where an item assessing distress was added after the introduction of DSM-IV) all of the analyses conducted above with and without an additional single-item measurement for distress added to the factor analysis from which we derived the syndromal impairment index.

When we compared the correlations between syndromal impairment ($n=788$) with and without the distress measure to our other two measures of depressive severity, the correlations rose slightly with criteria count (from 0.23 to 0.27) and with symptom severity (from 0.37 to 0.43). The strength of association of our measure of syndromal impairment to our wide range of validators also increased slightly with a mean (S.D.) of the ORs from 1.28 (0.25) to 1.33 (0.32), although the OR improved for only 14 of the 23 validators (for details see Table A1 in the Appendix). These results suggest a slight increase in the coherence and predictive power of the severity measures if distress is included in the measure of syndromal impairment. This comes, however, at the cost of a reduction in conceptual clarity as the constructs of syndrome-related distress and syndrome-related functional impairment are at least partially distinct.

Conclusions

Measures of the severity in MD are informative, telling us a range of useful things about expected patterns of co-morbidity, personality, clinical presentation and prognosis. Therefore, their inclusion as an 'episode specifier' for MD in DSM-IV makes clinical sense. However, the three specific measures of depressive severity included in DSM-IV (criteria count, syndromal impairment and symptom severity) are not equivalent. These three measures cannot be represented well by one or two of the other indices. Furthermore, although a general severity factor can be formed from these three measures, they do not, taken together, assess a single clear construct. Indeed, what is probably the most commonly used such measure, 'criteria count', in fact contributed the least to this general factor.

Our work supports the value of a clinical specifier of severity for MD and would argue for its inclusion in DSM-V. If the current clinical approach is adopted, the text should more clearly articulate the 'loose' or 'fuzzy' nature of the severity construct. Clinicians should, we suggest, be encouraged to average over the domains of criteria count, syndromal impairment and symptom severity, as dropping any one of them will result in a loss of information. Alternatively, further effort could be made to develop a specific scale to assess severity in MD as classified in the DSM. An empirically validated severity measure based on the DSM criteria would not add an important element to the clinical evaluation but could benefit clinical trials addressing treatment and interventions for different severity subtypes of depression. More detailed measurements, especially across a range of samples, might allow for superior predictive power and greater clarification of the structure of the severity of MD.

Appendix

Table A1. *Comparison of syndromal impairment with and without distress*

Dependent variable	Logit model	*n*	Syndromal impairment without distress OR (95% CI)	Syndromal impairment with distress OR (95% CI)
Co-morbidities				
General anxiety disorder (past year)	BL	788	1.05 (0.67–1.64)	1.03 (0.65–1.64)
Panic disorder	BL	654	1.71* (1.07–2.73)	1.72* (1.07–2.74)
Any phobia	BL	655	1.52*** (1.25–1.85)	1.60*** (1.30–1.97)
Alcohol dependence	BL	654	0.92 (0.75–1.11)	1.00 (0.82–1.23)
Nicotine dependence	BL	447	1.22 (0.96–1.55)	1.26 (0.98–1.62)
Any illicit drug dependence	BL	666	1.10 (0.84–1.45)	1.17 (0.88–1.55)
Adult antisocial personality traits	BL	618	1.23 (0.96–1.57)	1.26 (0.97–1.63)
Personality traits				
Higher neuroticism score	CL	782	1.75*** (1.41–2.16)	1.95*** (1.57–2.43)
Lower extraversion score	CL	782	1.28** (1.10–1.49)	1.28** (1.09–1.50)
Characteristics of index depressive episode				
Prominent anxiety symptoms co-occurring	BL	766	1.47** (1.12–1.92)	1.54** (1.16–2.05)
Duration of depressive episode in weeks	CL	788	1.53*** (1.24–1.89)	1.71*** (1.38–2.13)
Chronic MD	BL	788	1.75*** (1.39–2.19)	1.94*** (1.53–2.45)
'Out of the blue'	BL	788	1.27 (0.98–1.64)	1.28 (0.98–1.67)
Seeking help	BL	787	1.55*** (1.29–1.86)	1.75*** (1.44–2.12)
Prior depression history				
(Younger) Age of first onset of MD	CL	552	0.99 (0.80–1.22)	0.96 (0.77–1.20)
Number of episodes before index episode	CL	548	1.00 (0.83–1.20)	1.00 (0.83–1.21)
Demographic characteristics at index episode				
Age (being younger)	CL	788	1.23** (1.06–1.44)	1.28** (1.08–1.50)
Sex (being male)	BL	788	1.01 (0.85–1.21)	1.00 (0.83–1.20)
Not married or living with partner	BL	788	0.94 (0.79–1.12)	0.92 (0.76–1.10)
Lower family income	CL	788	1.07 (0.91–1.25)	1.05 (0.89–1.23)
Less school years	CL	788	0.94 (0.80–1.11)	0.97 (0.82–1.15)
Other validators				
New episode of MD in second wave	BL	660	1.36** (1.11–1.65)	1.50** (1.22–1.84)
MD diagnosis in co-twin[a]	BL	607	1.08 (0.86–1.35)	1.04 (0.84–1.30)

MD, Major depression; BL, binary logit; CL, cumulative logit; covariates: age, sex; [a] one-tailed *p* values, additional covariate: zygosity. *n*, sample size (differences due to missing data).

For the CL models, the ORs for a 1 standard deviation (S.D.) increase in the dependent variable are presented.

*** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$: two-tailed *p* values.

Declaration of Interest

None.

References

Alexopoulos GS, Bruce ML, Hull J, Sirey JA, Kakuma T (1999). Clinical determinants of suicidal ideation and behavior in geriatric depression. *Archives of General Psychiatry* **56**, 1048–1053.

Angst J, Amrein R, Stabl M (1995). Moclobemide and tricyclic antidepressants in severe depression: meta-analysis and prospective studies. *Journal of Clinical Psychopharmacology* **15**, S16–S23.

APA (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn. American Psychiatric Association: Washington, DC.

APA (1987). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn revised. American Psychiatric Association: Washington, DC.

APA (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association: Washington, DC.

Beck AT, Steer RA, Ball R, Ranieri W (1996). Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *Journal of Personality Assessment* **67**, 588–597.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.

Bentler PM (1990). Comparative fit indexes in structural models. *Psychological Bulletin* **107**, 238–246.

Blom MB, Spinhoven P, Hoffman T, Jonker K, Hoencamp E, Haffmans PM, van Dyke R (2007). Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression. *Journal of Affective Disorders* **104**, 119–126.

Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, Woo A, Trivedi MH (2006). The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *European Neuropsychopharmacology* **16**, 601–611.

Eysenck SBG, Eysenck HJ, Barrett P (1985). A revised version of the psychoticism scale. *Personality and Individual Differences* **6**, 21–29.

Fagerström KO (1978). Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addictive Behaviors* **3**, 235–241.

Faravelli C, Servi P, Arends JA, Strik WK (1996). Number of symptoms, quantification, and qualification of depression. *Comprehensive Psychiatry* **37**, 307–315.

Foley DL, Neale MC, Gardner C, Pickles A, Kendler KS (2003). Major depression and associated impairment: same or different genetic and environmental risk factors? *American Journal of Psychiatry* **160**, 2128–2133.

Gibbons RD, Clark DC, Kupfer DJ (1993). Exactly what does the Hamilton Depression Rating Scale measure? *Journal of Psychiatric Research* **27**, 259–273.

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

Hamilton MA (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology* **6**, 278–296.

Hiroe T, Kojima M, Yamamoto I, Nojima S, Kinoshita Y, Hashimoto N, Watanabe N, Maeda T, Furukawa TA (2005). Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. *Psychiatric Research* **135**, 229–235.

Hirschfeld RM (1999). Efficacy of SSRIs and newer antidepressants in severe depression: comparison with TCAs. *Journal of Clinical Psychiatry* **60**, 326–335.

Huang FY, Chung H, Kroenke K, Spitzer RL (2006). Racial and ethnic differences in the relationship between depression severity and functional status. *Psychiatric Services* **57**, 498–503.

Iannuzzo RW, Jaeger J, Goldberg JF, Kafantaris V, Sublette ME (2006). Development and reliability of the HAM-D/MADRS interview: an integrated depression symptom rating scale. *Psychiatric Research* **145**, 21–37.

Kasper S, Zivkov M, Roes KC, Pols AG (1997). Pharmacological treatment of severely depressed patients: a meta-analysis comparing efficacy of mirtazapine and amitriptyline. *European Neuropsychopharmacology* **7**, 115–124.

Kendler KS, Prescott CA (1999). A population-based twin study of lifetime major depression in men and women. *Archives of General Psychiatry* **56**, 39–44.

Kendler KS, Prescott CA (2006). *Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders*. Guilford Press: New York.

Kennedy N, Abbott R, Paykel ES (2004). Longitudinal syndromal and sub-syndromal symptoms after severe depression: 10-year follow-up study. *British Journal of Psychiatry* **184**, 330–336.

Kitamura T, Nakagawa Y, Machizawa S (1993). Grading depression severity by symptom scores: is it a valid method for subclassifying depressive disorders? *Comprehensive Psychiatry* **34**, 280–283.

Kroenke K, Spitzer RL, Williams JBW (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* **16**, 606–613.

Licht RW, Qvitzau S, Allerup P, Bech P (2005). Validation of the Bech–Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? *Acta Psychiatrica Scandinavica* **111**, 144–149.

Melartin TK, Rytsala HJ, Leskela US, Lestela-Mielonen PS, Sokero TP, Isometsa ET (2004). Severity and comorbidity predict episode duration and recurrence of DSM-IV major depressive disorder. *Journal of Clinical Psychiatry* **65**, 810–819.

Mojtabai R (2001). Impairment in major depression: implications for diagnosis. *Comprehensive Psychiatry* **42**, 206–212.

Montgomery SA, Äsberg M (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* **134**, 382–389.

Muthen LK, Muthen BO (2004). *Mplus User's Guide*. Muthen & Muthen: Los Angeles, CA.

Olsen LR, Jensen DV, Noerholm V, Martiny K, Bech P (2003). The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. *Psychological Medicine* **33**, 351–356.

Pettit JW, Lewinsohn PM, Roberts RE, Seeley JR, Monteith L (2009). The long-term course of depression: development of an empirical index and identification of early adult outcomes. *Psychological Medicine* **39**, 403–412.

Rodriguez BF, Bruce SE, Pagano ME, Keller MB (2005). Relationships among psychosocial functioning, diagnostic comorbidity, and the recurrence of generalized anxiety

disorder, panic disorder, and major depression. *Journal of Anxiety Disorders* **19**, 752–766.

Roy-Byrne PP, Stang P, Wittchen HU, Ustun B, Walters EE, Kessler RC (2000). Lifetime panic-depression comorbidity in the National Comorbidity Survey. Association with symptoms, impairment, course and help-seeking. *British Journal of Psychiatry* **176**, 229–235.

Rubenstein LV, Rayburn NR, Keeler EB, Ford DE, Rost KM, Sherbourne CD (2007). Predicting outcomes of primary care patients with major depression: development of a depression prognosis index. *Psychiatric Services* **58**, 1049–1056.

Santor DA, Coyne JC (2001). Examining symptom expression as a function of symptom severity: item performance on the Hamilton Rating Scale for Depression. *Psychological Assessment* **13**, 127–139.

SAS Institute (2005). *SAS OnlineDoc Version 9.1.3*. SAS Institute Inc.: Cary, NC.

Scheibe S, Preuschhof C, Cristi C, Bagby RM (2003). Are there gender differences in major depression and its response to antidepressants? *Journal of Affective Disorders* **75**, 223–235.

Sheehan DV, Harnett-Sheehan K, Raj BA (1996). The measurement of disability. *International Clinical Psychopharmacology* **11** (Suppl. 3), 89–95.

Strauss JS, Carpenter Jr. WT (1978). The prognosis of schizophrenia: rationale for a multidimensional concept. *Schizophrenia Bulletin* **4**, 56–67.

WHO (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization: Geneva.

WHO (1993). *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. World Health Organization: Geneva.

Zimmerman M, Ruggero CJ, Chelminski I, Young D, Posternak MA, Friedman M, Boerescu D, Attiullah N (2006). Developing brief scales for use in clinical practice: the reliability and validity of single-item self-report measures of depression symptom severity, psychosocial impairment due to depression, and quality of life. *Journal of Clinical Psychiatry* **67**, 1536–1541.

Zisook S, Rush AJ, Albala A, Alpert J, Balasubramani GK, Fava M, Husain M, Sackeim H, Trivedi M, Wisniewski S (2004). Factors that differentiate early vs. later onset of major depression disorder. *Psychiatric Research* **129**, 127–140.

Zung WW (1965). A Self-Rating Depression Scale. *Archives of General Psychiatry* **12**, 63–70.