

Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria

V. Lux¹ and K. S. Kendler^{2*}

¹ Department of Psychology, Free University Berlin, Germany

² Virginia Institute for Psychiatric and Behavioral Genetics and Departments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA USA

Background. The DSM-IV symptomatic criteria for major depression (MD) derive primarily from clinical experience with modest empirical support.

Method. The sample studied included 1015 (518 males, 497 females) Caucasian twins from a population-based registry who met criteria for MD in the year prior to the interview. Logistic regression analyses were conducted to compare the associations of: (1) single symptomatic criterion, (2) two groups of criteria reflecting cognitive and neurovegetative symptoms, 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. The individual symptomatic criteria showed widely varying associations with the pattern of co-morbidity, personality traits, features of the depressive episode and demographic characteristics. When examined separately, these two criteria groups showed robust differences in their patterns of association, with the validators with the cognitive criteria generally producing stronger associations than the neurovegetative.

Conclusions. Among depressed individuals, individual DSM-IV symptomatic criteria differ substantially in their predictive relationship with a range of clinical validators. These results challenge the equivalence assumption for the symptomatic criteria for MD and suggest a more than expected degree of ‘covert’ heterogeneity among these criteria. Part of this heterogeneity is captured by the distinction between cognitive *versus* neurovegetative symptoms, with cognitive symptoms being more strongly associated with most clinically relevant characteristics. Detailed psychometric evaluation of DSM-IV criteria is overdue.

Received 30 June 2009; Revised 5 November 2009; Accepted 17 November 2009; First published online 11 January 2010

Key words: Criteria symptoms, DSM-IV, heterogeneity, major depression, psychiatric diagnosis.

Introduction

A key component of the current approach to the diagnosis of major depression (MD) in DSM-IV is the symptomatic or ‘A’ criteria. These nine criteria (‘depressed mood’, ‘loss of interest’, ‘appetite or weight change’, ‘hyper- or insomnia’, ‘psychomotor agitation or retardation’, ‘fatigue’, ‘feelings of worthlessness or guilt’, ‘diminished ability to think or concentrate or indecisiveness’ and ‘suicidal ideation or thoughts of death’) derive, with only modest alteration, from the list of symptoms proposed in the Feighner *et al.* (1972) criteria. These criteria were created by consensus of senior clinicians at Washington University in the early 1970s and were particularly influenced by a study by Cassidy *et al.*

(1957) that analyzed symptom frequencies in manic-depressive patients as compared with those with medical illnesses (Cassidy *et al.* 1957; Zimmerman *et al.* 2006b; Kendler *et al.* 2009). The symptoms included in the Cassidy study were, in turn, partly based on those utilized in a 1950 report, which examined symptoms in 50 cases of melancholia (Stone & Burris, 1950).

Since then, many studies have examined the diagnostic reliability of this classification (for example, Spitzer & Fleiss, 1974; Spitzer *et al.* 1978; Keller *et al.* 1995), the differentiation between clinical and sub-clinical depression (for example, Kendler & Gardner, 1998; Parker *et al.* 1998), the validity of subtypes (Frances *et al.* 1990; Kendler *et al.* 1996; Schotte *et al.* 1997; Chen *et al.* 2000; Sullivan *et al.* 2002; Carragher *et al.* 2009) and the relationship between single symptoms and depressive subtypes with severity of MD (Faravelli *et al.* 1996). By contrast, we are aware of only two studies that have addressed, even indirectly, the validity of these widely used symptomatic criteria

* 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)

for MD (McGlinchey *et al.* 2006; Zimmerman *et al.* 2006a). Both concluded that further efforts at validation were needed. Zimmerman *et al.* (2006a) examined the content validity of the DSM-IV classification of MD and found the criteria were somewhat redundant. The same research group examined the ability of single depressive symptoms to differentiate between patients with and without a diagnosis of MD and found that the individual criterion performed quite differently in this regard (McGlinchey *et al.* 2006).

In this report, we examine the validity of the DSM-IV symptomatic criteria for MD in a sample of ~1000 cases of MD reported in the last year in a population-based cohort of twins (Kendler & Prescott, 2006). First, we examine the ability of individual criterion to predict uniquely a broad set of validators, including risk of MD in co-twin, demographic characteristics, lifetime co-morbidities, characteristics at index episode, prior depression history and risk for future episodes. Based on these results, we then aggregate two sets of symptomatic MD criteria 'A' and compare their performance on the same set of validators.

Methods

Sample and measurements

Participants in this report derive from two inter-related studies in Caucasian same-sex twin pairs from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD) (Kendler & Prescott, 2006). All subjects 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, the response rate to which was ~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%. Participating twin pairs were identified for the male–male and male–female (MMMMF) cohort from twins with birth years from 1940–1974 initially ascertained directly from registry records. The first MMMF interview (MMMMF1) was conducted by telephone with a response rate of 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 496 pairs (Kendler & Prescott, 1999).

MD diagnoses were based on a module in the FF1 and MMMF1, where every subject was asked whether they experienced each of the disaggregated criteria

symptoms for DSM-IV MD in the year prior to the interview. Separate questions were asked for psychomotor agitation and retardation, insomnia and hypersomnia, weight loss and gain and appetite increase and decrease.

DSM-IV criteria for last-year MD were met by 217 twins from the FF1 and 798 twins from the MMMF1 interviews, prevalence rates are 10% [95% confident interval (CI) 8.8–11.4%] in the FF1 and 11.7% (95% CI 11.0–12.5%) in the first interview of male–female twins. Of these 1015 twins, 518 were males and 497 were females. At the time of the first interview, their ages ranged from 18 to 57 years with a mean of 34.5. There were 83 twin pairs, with both twins diagnosed with MD, and two more pairs from two triplets who were diagnosed with MD.

For our first analysis, we created binary dummy variables for each of the nine criteria, indicating the presence of each as part of the depressive syndrome. For the disaggregated criteria, it was counted present when at least one of the disaggregated symptoms was reported (e.g. 'weight loss' for appetite/weight change). In addition, a criterion was not counted as part of the depressive syndrome when reported due to medication or illness.

For comparing the two symptom groups – cognitive and neurovegetative – we counted the number of endorsed symptoms separately for each group. Cognitive symptoms included 'depressed mood', 'loss of interest', 'worthlessness/guilt' and 'suicidal ideation' and neurovegetative symptoms included 'sleep', 'appetite/weight', 'psychomotor' changes and 'fatigue'. 'Trouble concentrating' was excluded from this analysis (see Results for details).

The VATSPSUD includes a rich set of data about future episodes, co-twin history of MD, 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 future episodes, depressive episodes of the co-twin and all other co-morbidities, data were obtained from all interview waves to maximize available information.

GAD was diagnosed using the DSM-III-R criteria (APA, 1987) requiring a minimum of 1 month duration. Panic disorder was also diagnosed using the DSM-III-R criteria. We also examined a broad definition of panic ('panic broad'), for which the only criterion was a positive response to a probe question for lifetime panic attacks ('Thinking back over your entire life, have you ever had a spell or attack when you suddenly felt frightened or extremely

uncomfortable in a situation in which you didn't expect to feel that way?'). Real danger or clear phobic stimuli were excluded. '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 ≥ 7 on the Fagerström Tolerance Questionnaire (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 three or more 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 for which we had separate questions: 'felt anxious, nervous or worried' and 'muscles felt tense or felt jumpy or shaky inside' lasting at least 5 days in the last year prior to the interview. 'Chronic MD' was defined as a depressive episode lasting 12 months or longer. For 'something happened before the depressive episode', we used a question asking whether 'something happened to make you feel that way or did the feeling just come on you "out of the blue"?' for the worst depressive episode in the last year. '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. Finally, childhood sexual abuse was defined by a positive response to the question: 'Have you ever been sexually abused or molested?' and when the abuse happened before the age of 16 years.

Data analysis

We conducted logistic regression analyses with the LOGISTIC function in SAS (SAS Institute, 2005). Our approach was to compare the odds ratios (ORs) of each test to examine possible patterns of differences, with the p values indicating the general probability of results arising from chance. (For our approach to the question of multiple testing, see Limitations.) For the first set of analyses, the validator variable functioned as dependent variable while the dummy variable representing that the specific symptom was part of the syndrome was the predictor variable. For the second set of analyses, which compared the two symptom groups, the number of cognitive or number of neurovegetative symptoms was the predictor variable. Depending on the validator variable, we fitted a

binary or cumulative logit model to the data, with age and sex as covariates. In addition, we included the number of positive criteria symptoms as a covariate in the model when comparing the single symptoms with control for the overall clinical severity of the depressive episode as indexed by the number of endorsed criteria. When we examined the two symptom groups, the symptom counts for both groups were included in the model to get the unique predictive power for the cognitive and the neurovegetative symptom count controlling for the number of symptoms endorsed in the other symptom group. In addition, we also tested if the observed differences in the odds ratios between both groups were significant using the TEST statement of the LOGISTIC procedure in SAS (SAS Institute, 2005).

When risk of MD in co-twin was the dependent variable, zygosity was included as covariate in the model. 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.

In interpreting these results, it is important to note that the first set of analyses that examined individual criterion controlled for total number of endorsed symptoms. In these analyses, an OR < 1 would indicate that compared with all subjects who endorsed n criteria, those who endorsed symptom X predicted the validator more poorly. However, in our analyses of criterion groups, controlling for number of endorsed symptoms had the undesirable effect of causing a strong negative correlation between the predictive effects of the two symptom groups. Therefore, in these analyses, where we did not control for total endorsed criteria, we would expect nearly all ORs to be positive as more symptoms tend to predict these validators. Our goal here was to compare the strength of the observed positive ORs between the two symptom groups.

Results

Logistic regression analyses of single symptoms

The first nine columns of Table 1 show the results from the logistic regression analyses of the nine A criteria with each of the 25 validator variables. ORs, 95% CI and levels of statistical significance are presented, controlling for age, sex and the total number of endorsed criteria. For the cumulative logit models, the ORs reflect the impact of a 1 s.d. change in the dependent variable.

In total, 36 of the 225 analyses (16%) were statistically significant ($p < 0.05$), far in excess of chance expectations (Feild & Armenakis, 1974). Loss of

Table 1. Comparison major depression (MD) criteria symptoms

Response variable	<i>n</i>		A1 Depressed mood	A2 Loss of interest	A3 Appetite/ Weight	A4 Sleep	A5 Psychomotor Agitation/ Retardation	A6 Fatigue	A7 Worthless- ness/Guilt	A8 Trouble concentrating	A9 Suicidal ideation
Co-morbidities lifetime											
General anxiety disorder (last year)	1015	OR	1.11	0.80	0.95	0.95	2.02***	0.72	0.79	1.15	1.11
BL		CI	0.23–5.34	0.44–1.44	0.59–1.51	0.58–1.56	1.19–3.45	0.46–1.12	0.52–1.20	0.75–1.76	0.70–1.75
Panic (broad)	859	OR	1.20	1.17	0.79	0.66*	1.45	0.93	1.10	0.92	1.28
BL		CI	0.27–5.38	0.60–2.26	0.51–1.23	0.41–1.06	0.85–2.49	0.59–1.49	0.72–1.67	0.61–1.40	0.84–1.94
Panic disorder	859	OR	xxx	0.58	0.69	0.60	1.10	1.08	1.46	0.73	2.02**
BL		CI		0.22–1.50	0.33–1.44	0.28–1.27	0.45–2.67	0.48–2.46	0.71–3.00	0.37–1.45	1.04–3.95
Any phobia	881	OR	1.19	0.76	1.00	0.76	1.30	0.97	1.35*	0.75*	1.18
BL		CI	0.45–3.13	0.49–1.17	0.72–1.39	0.53–1.08	0.91–1.86	0.70–1.35	1.00–1.84	0.55–1.02	0.85–1.65
Alcohol dependence	880	OR	1.85	0.97	0.81	0.83	0.98	0.98	1.48**	0.87	1.07
BL		CI	0.64–5.33	0.61–1.55	0.58–1.15	0.57–1.20	0.67–1.44	0.69–1.40	1.06–2.06	0.63–1.20	0.76–1.52
Nicotine dependence	558	OR	3.54	1.03	0.90	0.72	1.35	0.62**	1.60**	0.76	1.25
BL		CI	0.71–17.62	0.57–1.83	0.59–1.39	0.46–1.14	0.85–2.15	0.40–0.96	1.07–2.38	0.51–1.13	0.83–1.90
Any illicit drug dependence	844	OR	1.81	0.91	0.54***	0.89	1.22	0.76	1.21	1.02	1.72**
BL		CI	0.40–8.09	0.48–1.75	0.34–0.86	0.53–1.50	0.71–2.10	0.48–1.22	0.76–1.92	0.65–1.59	1.10–2.70
Adult antisocial personality trait	753	OR	1.35	1.01	0.58**	0.91	1.01	0.64*	1.18	1.06	1.97***
BL		CI	0.37–4.86	0.54–1.90	0.37–0.91	0.55–1.50	0.60–1.69	0.41–1.01	0.76–1.84	0.69–1.64	1.29–3.02
Personality traits											
Higher neuroticism score	977	OR	1.37	0.72*	0.67***	0.66***	0.93	1.08	2.06‡	0.77*	1.48***
CL		CI	0.56–3.34	0.50–1.05	0.51–0.90	0.49–0.90	0.68–1.26	0.82–1.43	1.58–2.69	0.59–1.01	1.11–1.98
Higher extraversion score	977	OR	0.72	1.15	1.16	1.21	1.11	1.02	0.67***	1.25*	0.77*
CL		CI	0.31–1.67	0.81–1.65	0.89–1.52	0.90–1.61	0.83–1.48	0.78–1.33	0.52–0.87	0.97–1.60	0.59–1.01
Characteristics of index episode											
Anxiety symptoms co-occurring	992	OR	0.73	0.79	0.48***	1.38	2.04†	0.71	1.37	1.03	0.79
BL		CI	0.21–2.62	0.45–1.39	0.30–0.77	0.89–2.12	1.33–3.11	0.46–1.10	0.91–2.06	0.68–1.55	0.48–1.29
Duration depressive episode in weeks	1015	OR	0.78	0.66	0.92	0.73	0.77	0.66**	1.27	1.03	2.13‡
CL		CI	0.21–2.88	0.38–1.15	0.61–1.40	0.48–1.13	0.48–1.22	0.44–0.98	0.86–1.88	0.70–1.51	1.47–3.09
Chronic MD (duration ≥ 12 months)	1015	OR	0.61	0.62	0.98	0.76	0.80	0.64**	1.27	1.03	2.06†
BL		CI	0.17–2.23	0.34–1.10	0.63–1.54	0.48–1.20	0.49–1.33	0.42–0.98	0.84–1.93	0.69–1.56	1.39–3.06

Something happened before episode	1015	OR	0.80	1.88**	1.52*	1.16	0.97	0.69	1.13	1.33	0.42‡
BL		CI	0.18–3.56	1.07–3.30	0.97–2.38	0.70–1.90	0.57–1.63	0.42–1.13	0.73–1.75	0.86–2.05	0.28–0.65
Seeking help	1014	OR	4.18*	0.77	1.01	0.98	0.72*	0.76*	1.16	0.83	1.91‡
BL		CI	0.95–18.34	0.51–1.18	0.74–1.39	0.70–1.38	0.51–1.01	0.56–1.04	0.86–1.56	0.61–1.11	1.39–2.61
Prior depression history											
Number of episodes at 1 st interview	715	OR	0.72	1.07	0.92	0.90	1.00	1.19	1.06	0.98	0.96
CL		CI	0.26–1.98	0.68–1.67	0.67–1.27	0.64–1.26	0.70–1.43	0.87–1.64	0.78–1.42	0.72–1.32	0.70–1.32
Age of first onset of MD	719	OR	0.65	0.82	0.85	1.72**	1.06	0.92	0.92	1.12	0.87
CL		CI	0.18–2.26	0.48–1.38	0.58–1.23	1.13–2.62	0.68–1.63	0.63–1.34	0.65–1.31	0.78–1.61	0.59–1.28
Demographic characteristics at index episode											
Age (being older)	1015	OR	5.49***	1.14	0.94	0.75*	1.45**	1.05	0.78*	0.89	1.12
CL		CI	1.98–15.28	0.79–1.65	0.71–1.23	0.56–1.01	1.08–1.96	0.80–1.39	0.60–1.01	0.68–1.15	0.85–1.49
Sex (being female)	1015	OR	3.46**	1.05	1.55***	1.05	0.71**	1.48**	0.85	0.79*	0.75*
BL		CI	1.24–9.68	0.70–1.57	1.14–2.10	0.76–1.45	0.51–0.98	1.10–2.00	0.64–1.12	0.59–1.04	0.55–1.02
Married or living with partner	1015	OR	0.45*	0.84	1.04	0.82	1.27	1.08	1.03	0.87	1.19
BL		CI	0.17–1.15	0.56–1.28	0.76–1.41	0.59–1.14	0.91–1.76	0.80–1.47	0.77–1.37	0.65–1.16	0.87–1.64
More school years	1015	OR	0.78	1.21	1.02	1.40**	0.71**	1.32**	0.96	1.01	0.71**
CL		CI	0.33–1.86	0.83–1.77	0.77–1.36	1.04–1.89	0.52–0.97	1.00–1.75	0.73–1.25	0.77–1.31	0.53–0.94
Higher family income	1015	OR	0.25***	1.03	0.81	1.33*	0.96	1.25	0.88	1.33**	0.75
CL		CI	0.09–0.66	0.71–1.50	0.61–1.07	0.98–1.77	0.71–1.30	0.95–1.64	0.68–1.14	1.03–1.73	0.57–1.00
Other validators											
Childhood sexual abuse	900	OR	0.76	1.08	0.97	0.87	1.49	0.90	1.05	0.68*	1.32
BL		CI	0.20–2.89	0.59–1.97	0.61–1.55	0.53–1.41	0.90–2.48	0.57–1.42	0.69–1.60	0.45–1.04	0.85–2.05
New episode of MD	876	OR	0.57	1.55*	0.91	1.01	1.09	1.03	0.98	0.95	0.92
BL		CI	0.22–1.46	0.96–2.49	0.65–1.27	0.70–1.45	0.75–1.56	0.74–1.45	0.72–1.34	0.69–1.30	0.65–1.28
MD diagnosis in co-twin ^a	800	OR	0.55	1.04	1.05	1.20	1.03	1.14	1.04	0.92	0.76
BL		CI	0.21–1.46	0.66–1.63	0.75–1.47	0.83–1.74	0.71–1.49	0.82–1.60	0.76–1.43	0.67–1.26	0.53–1.07
Mean OR (s.d.)			1.99 (1.18)	1.25 (0.24)	1.26 (0.29)	1.28 (0.19)	1.28 (0.29)	1.25 (0.20)	1.25 (0.23)	1.17 (0.14)	1.46 (0.39)

^a One-tailed *p* values, regression model with zygosity as additional covariate;

n, sample size, differences in sample sizes throughout validators due to missing data; BL, binary logit model; CL, cumulative logit model, covariates included in the model: age, sex and criteria symptom count; xxx, no valid regression model; OR, odds ratio; CI, confidence interval.

* *p* < 0.1, shaded cells are significant at the *p* < 0.05 level, ** *p* < 0.05, *** *p* < 0.01.

† *p* < 0.001, ‡ *p* < 0.0001.

interest was the symptom with the fewest (only one) and suicidal ideation was the symptom with the most significant results (nine). The means (s.d.) of the ORs for the nine criteria across all the analyses ranged from 1.99 (1.18) for depressed mood to 1.17 (0.14) for trouble concentrating.

These results reveal a pattern of relationships between the criteria and the validators that was unexpectedly complex and variable. Of the diverse set of findings, nine were noteworthy.

- (1) Regarding co-morbidities with anxiety disorders, psychomotor agitation/retardation predicted co-morbidity with GAD and was strongly positive associated with prominent anxiety symptoms, whereas appetite/weight change was negative associated with the anxiety symptoms. In addition, suicidal ideation positively predicted fully syndromal panic disorder.
- (2) With regard to co-morbidities with substance use disorders and adult antisocial personality traits, suicidal ideation positively predicted risk for illicit drug dependence and adult antisocial personality traits, while appetite or weight changes were negative associated with both these validator variables. Yet, only worthlessness/guilt predicted co-morbidity with alcohol dependence and nicotine dependence, while fatigue was negative associated with nicotine dependence.
- (3) None of the nine A criteria was significantly related to a new episode of MD or a diagnosis of MD in the co-twin. However, examining non-significant findings ($p > 0.05$), we observed a heterogeneous pattern of relationships for these validators. For example, loss of interest showed a positive but depressed mood a negative association with risk for future episodes. Problems sleeping and fatigue tended to be positively related and depressed mood and suicidal ideation negatively related to the risk for MD in the co-twin.
- (4) Significant sex differences emerged for four of the criteria; depressed mood, appetite/weight changes and fatigue significantly predicted being female, while psychomotor agitation/retardation was associated with being male.
- (5) The criteria varied substantially in their relationships to other demographic characteristics. For example, age differences were found for two criteria; depressed mood and psychomotor agitation/retardation had a significant positive correlation with current age. In addition, four symptoms were associated with education, with sleep changes and fatigue being associated with more, and psychomotor agitation/retardation and suicidal ideation associated with less years of

education. Also, depressed mood was associated with lower and trouble concentrating associated with higher family income.

- (6) Regarding the two personality traits – neuroticism and extraversion – feeling worthless/guilty and suicidal ideation predicted higher neuroticism scores, while appetite/weight change and sleep changes were associated with lower neuroticism scores. In addition, worthlessness/guilt was associated with a lower extraversion score.
- (7) Only suicidal ideation predicted a longer duration of the index episode and chronic depression, while fatigue was positively associated with shorter episodes and negatively associated with ‘chronic depression’ (defined as lasting ≥ 12 months).
- (8) Loss of interest was associated with episodes precipitated by a particular life event (‘something happened’), while suicidal ideation was associated with experiencing the depressive episode ‘out of the blue’.
- (9) None of the symptoms was significantly associated with the number of prior depressive episodes, while only sleep problems predicted (older) age of first onset of MD.

Logistic regression analyses of the two symptom groups

A review of the results obtained at the individual symptom level reveals a tendency for predictions from neurovegetative symptoms (most typically appetite/weight, sleep problems, psychomotor changes and fatigue) to differ qualitatively from those obtained for the ‘cognitive/emotional’ symptoms (hereafter ‘cognitive’) of sad mood, loss of interest, guilt and suicidal ideation. For example, a significantly elevated risk for alcohol, nicotine and illicit drug dependence and antisocial traits is associated with symptoms of guilt or suicidal ideation, whereas a significantly decreased risk is associated with symptoms of fatigue or appetite changes. Levels of neuroticism are positively associated with guilt or suicidal ideation but negatively associated with appetite and sleep changes. Fatigue predicts shorter and suicidal ideation longer depressive episodes. Similar trends that do not reach significance are seen for co-morbidity, with panic disorder (positive with suicidal ideation and negative with appetite and sleep problems) and alcoholism (positive with guilt and negative with appetite and sleep problems), help-seeking (positive with suicidal ideation and negative with psychomotor changes and fatigue) and age at onset (positive with sleep problems and negative with depressed mood, loss of interest and suicidal ideation). While this trend was far from uniform (and did not appear to involve

trouble concentrating in either group), we judged it sufficiently prominent to warrant further investigation.

We therefore examined, in our sample of individuals with MD in the last year, the ability of the sum of the symptoms endorsed in these two groups to predict the same set of validators we examined for the individual symptoms. Table 2 shows the results of these analyses.

ORs, 95% CI and levels of statistical significance are presented, controlling only for age, sex and the number of endorsed criteria symptoms of the other symptom group. We also controlled for zygosity when the 'MD diagnosis in co-twin' was the dependent variable. For the cumulative logit models, the ORs of 1 s.d. increase of the dependent variable are reported in the table.

Strikingly, for 17 of the 25 validator variables, the predictive power of the two symptom groups was significantly different from one another. Out of these results, eight were noteworthy.

- (1) The number of endorsed cognitive criteria better predicted co-morbidities with other disorders than did the number of neurovegetative criteria, and these differences were significant except for GAD and our broad definition of panic.
- (2) The number of cognitive criteria better predicted both higher neuroticism scores and lower extraversion scores and these differences were also significant.
- (3) Regarding characteristics of the index episode, the cognitive criteria count significantly better predicted the duration of the episode, chronic MD defined as a depressive episode lasting longer than 12 months and seeking help, while the number of neurovegetative criteria better predicted that the depressive episode occurred 'out of the blue'.
- (4) The cognitive symptom count also significantly better predicted the number of prior episodes.
- (5) The two symptom scores did not predict age at interview while the neurovegetative criteria count better predicted being female; this difference was significant.
- (6) Regarding the other demographic characteristics, the cognitive criteria count was associated with fewer years of schooling and lower income; these differences were also significant.
- (7) The cognitive criteria count significantly better predicted childhood sexual abuse.
- (8) We found no significant differences between the two criteria counts regarding the prediction of a new depressive episode and a MD diagnosis in the co-twin.

Discussion

The aim of our analysis was to examine carefully the validity of the individual DSM-IV symptomatic or 'A' criteria for MD. In a large epidemiological sample of individuals who reported episodes of MD in the last year, we began by comparing the nine criteria on a wide range of potentially validating characteristics, none of which played any role in the diagnostic process. Our analyses focused on potential similarities and differences in the patterns of association of the nine criteria symptoms with these characteristics. In these analyses, we controlled for the number of endorsed A criteria so that an OR >1 meant that the presence of this criterion predicted the validator more strongly than expected for individuals with the same number of endorsed symptoms without that particular criterion. An OR <1 meant that the presence of this criterion predicted the validator more poorly than expected for individuals with the same number of endorsed symptoms without that particular criterion.

These results revealed an unexpected degree of heterogeneity in the performance of the individual criterion. Indeed, each of the nine criteria had a relatively unique pattern of findings with regard to our set of validator variables, suggesting a surprising degree of evidence for 'covert heterogeneity' within the MD syndrome. The relationships we observed support the notion of heterogeneity of the classification of MD discussed in the literature (e.g. Merikangas *et al.* 1994; Kendler *et al.* 1996; Chen *et al.* 2000; Carragher *et al.* 2009; Hybels *et al.* 2009).

To further characterize this heterogeneity, we also examined an intermediate level between the MD diagnosis and the single criterion by comparing the number of endorsed criteria separately for two symptom groups – cognitive and neurovegetative – through the same set of validator variables. Of note, in recent exploratory factor analyses of depressive symptoms run with the entire VATSPSUD sample, we found a good statistical fit for a two-factor model with one factor reflecting cognitive and the other neurovegetative symptoms. When we compared these two symptom groups, we did not control for the total number of endorsed symptoms. Rather, we specifically examined whether the predictive power differed between the two groups of criteria. As our results show, some of this heterogeneity can be represented by the differences between the number of endorsed cognitive and neurovegetative criteria, with these differences being significant for most of the validator variables. Therefore, co-morbidities with other psychiatric disorders and the personality traits – higher neuroticism and lower extraversion – were most

Table 2. Comparison of cognitive and neurovegetative criteria count (logistic regression)

Dependent variable	<i>n</i>		Cognitive criteria count (CCC)	Neurovegetative criteria count (NCC)	χ^2 test CCC=NCC (<i>p</i> values)
Co-morbidities (lifetime)					
General anxiety disorder (last year)	1015	OR	1.55‡	1.45‡	0.5111
BL		CI	1.39–1.72	1.32–1.60	
Panic (broad)	859	OR	1.35‡	1.20‡	0.1531
BL		CI	1.23–1.47	1.11–1.30	
Panic disorder	859	OR	1.60‡	1.13*	0.0173
BL		CI	1.36–1.87	0.98–1.31	
Any phobia	881	OR	1.30‡	1.16‡	0.0282
BL		CI	1.23–1.38	1.10–1.23	
Alcohol dependence	880	OR	1.36‡	1.17‡	0.0138
BL		CI	1.27–1.45	1.11–1.25	
Nicotine dependence	558	OR	1.26‡	1.09**	0.0481
BL		CI	1.16–1.36	1.02–1.18	
Any illicit drug dependence	844	OR	1.56‡	1.07	<0.0001
BL		CI	1.42–1.71	0.97–1.16	
Adult antisocial personality trait	753	OR	1.54‡	1.01	<0.0001
BL		CI	1.41–1.69	0.92–1.20	
Personality traits					
Higher neuroticism score	977	OR	2.15‡	1.41‡	<0.0001
CL		CI	1.94–2.37	1.29–1.54	
Higher extraversion score	977	OR	0.82‡	1.02	<0.0001
CL		CI	0.77–0.87	0.96–1.08	
Characteristics of index depressive episode					
Prominent anxiety symptoms	992	OR	1.57‡	1.64‡	0.4120
BL		CI	1.48–1.68	1.55–1.74	
Duration depressive episode in weeks	1015	OR	1.76‡	1.43‡	0.0260
CL		CI	1.46–1.68	1.29–1.46	
Chronic MD (duration \geq 12 months)	1015	OR	1.77‡	1.41‡	0.0259
BL		CI	1.56–2.01	1.26–1.59	
Something happened	1015	OR	1.02	0.89***	0.0475
BL		CI	0.94–1.11	0.83–0.97	
Seeking help	1014	OR	1.60‡	1.30‡	0.0004
BL		CI	1.49–1.73	1.21–1.39	
Prior depression history					
Number of episodes	715	OR	1.37‡	1.20‡	0.0139
CL		CI	1.29–1.46	1.13–1.27	
Age of first onset of MD > 18	719	OR	0.82‡	0.90***	0.1616
BL		CI	0.76–0.89	0.84–0.96	
Demographic characteristics at index episode					
Age (being older)	1015	OR	0.95**	1.03	0.0517
CL		CI	0.90–1.00	0.99–1.08	
Sex (being female)	1015	OR	1.02	1.16‡	0.0070
BL		CI	0.97–1.08	1.11–1.22	
Married or living with partner	1015	OR	0.86‡	0.92***	0.1435
BL		CI	0.81–0.91	0.88–0.97	
More than 9 school years (education)	1015	OR	0.77‡	0.99	0.0212
BL		CI	0.69–0.87	0.89–1.11	
Higher Family income	1015	OR	0.86‡	0.96*	0.0310
CL		CI	0.82–0.91	0.91–1.01	
Other validators					
Childhood sexual abuse (broad)	900	OR	1.31‡	1.12***	0.0488
BL		CI	1.20–1.43	1.04–1.21	

Table 2 (cont.)

Dependent variable	<i>n</i>		Cognitive criteria count (CCC)	Neurovegetative criteria count (NCC)	χ^2 test CCC=NCC (<i>p</i> values)
New episode of MD in second wave	876	OR	1.61‡	1.45‡	0.0955
BL		CI	1.50–1.73	1.36–1.55	
MD diagnosis in co-twin ^a	800	OR	1.64‡	1.48‡	0.1272
BL		CI	1.52–1.78	1.38–1.59	

n, Sample size, differences due to missing data; BL, binary logit model; CL, cumulative logit model, covariates included in the model: sex, age and symptom count of the other symptom group; MD, major depression; OR, odds ratio; CI, confidence interval.

Cognitive symptoms: depressed mood; loss of interest; feeling worthless/guilty; suicidal ideation.

Neurovegetative symptoms: appetite/weight change; sleep; psychomotor agitation/retardation; fatigue.

Not counted: trouble concentrating.

^a One-tailed *p* values, zygosity included as covariate in the model.

Two-tailed *p* values: * *p* < 0.1, ** *p* < 0.05, *** *p* < 0.01, † *p* < 0.001, ‡ *p* < 0.0001.

strongly associated with the cognitive criteria group, while we found significant sex differences (predicting being female) for the number of endorsed neurovegetative criteria.

Some of our individual findings have precedents in the literature. For example, Keller *et al.* (2007) found that thoughts of self-harm were not associated with adverse life events, which are partly reflected in our result of suicidal ideation, predicting the depressive episode as experienced 'out of the blue'. Although sex differences for psychomotor changes and fatigue have already been reported for our sample by Khan *et al.* (2002), our approach was probably not sensitive enough to also show the sleep differences that Khan *et al.* reported in their study of matched twins. Corresponding to our results, Angst and Dobler-Mikola (1984) found that females report weight change more often when diagnosed with MD. Their finding of females reporting more feelings of worthlessness/guilt is not supported by our results. In addition, other studies found no sex differences for expressed depressive symptoms (Middeldorp *et al.* 2006). Furthermore, some associations between substance use disorders and psychomotor agitation and appetite/weight loss were reported in the literature (for example, Balazs *et al.* 2006; Leventhal *et al.* 2008), but they are not comparable with our results because of the use of disaggregated symptoms in these studies.

The question of homogeneity *versus* heterogeneity of MD, including the distinction of 'types' of depression by symptom groups or causes (endogenous *versus* reactive) has a long history in psychiatry (for the homogeneity/unitary position, see Mapother, 1926; Lewis, 1934; Kendell, 1969; Akiskal & McKinney, 1975; for the heterogeneity/binary position, for

example, Gillespie, 1929; Kiloh & Garside, 1963; Eysenck, 1970). Because of a lack of empirical support for a clear binary concept of MD, DSM-III and IV definition were based on the unitarian position accompanied with the hope of finally settling these debates (see Parker, 2000; van Praag, 2000). However, the validity of the current DSM-IV classification of MD has been questioned (see e.g. Parker, 2005; Zimmerman *et al.* 2006*b*). The amount of heterogeneity that we found comparing the performance of criteria symptoms support these concerns. Yet, our results indicate that only some of this heterogeneity can be captured by a simple binary structure of cognitive *versus* neurovegetative symptoms.

Our results also have implications for the basic structure of DSM diagnostic categories. Beginning with DSM-III, 'polythetic' diagnostic models formed the basis for DSM diagnoses. Such models implicitly assume an approximate equivalence of the individual criterion – that is, that each criterion plays a similar and roughly interchangeable role in the diagnostic process. Our results question this assumption for the MD symptomatic criteria. In fact, we found that different symptoms or groups of symptoms of MD are associated with different clinical characteristics.

While each DSM revision has been justifiable proud about the increasing role of empirical findings in making nosologic decisions, in fact most of the diagnostic categories and the diagnostic criteria they contain have been accepted for historical rather than strictly empirical reasons (Kendler & Zachar, 2008). Given the deep clinical wisdom and experience contained in these traditions, this is not an illogical approach. However, these results provide some challenge to this approach and, at a minimum, suggest that

detailed psychometric evaluation of these historically defined diagnostic entities is overdue.

Finally, we find that cognitive–emotional criteria are rather consistently stronger predictors of our validators than are the neurovegetative criteria. In contrast, common clinical teaching highlights the neurovegetative or ‘biological’ symptoms as the ‘core’ features of depression. These results are broadly consistent with the work of Beck and colleagues, which emphasizes the primacy of cognitive changes in the etiology of MD (Beck & Alford, 2008).

Limitations

These results should be interpreted in the light of four 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 nature of our validator variables made it difficult to account formally, in most of our analyses, for the non-independence of observations in our twin data. However, only ~17% of our data result from twin pairs and by exploring formal corrections for the binary logit models we found no substantial effects. Third, given the number of tests we performed, some of the significant findings are surely a result of chance effects. To correct formally for multiple testing, a strict Bonferroni (see Abdi, 2007) would require setting the p value to around 0.0002. This would surely be too conservative, because of the substantial high correlations between a number of symptoms and between some of the validators. Furthermore, our goal here is not so much to strictly evaluate every single hypothesis; rather, we advocate taking an experimental wide approach, asking if the broad pattern of relationships between symptoms and validators are likely chance effects. As we demonstrate, this is very unlikely to be the case, as the total number of significant findings far exceeds that expected under the null hypothesis. While we cannot be confident that any one of our findings is not due to chance effects, the probability that the entire pattern of results arose solely from false positives due to multiple testing is extremely unlikely.

Fourth, we created the two symptom groups based partly on our initial results and partly on conceptual and clinical grounds. Only the decision to drop ‘trouble concentrating’ was based on empirical results from our single symptom analysis. Other clusters of symptom groups could have possibly revealed other structures. However, the distinction between cognitive and neurovegetative symptoms has long precedence in the depression literature (Koenig *et al.* 1993; Beck & Alford, 2008; Carragher *et al.* 2009) and

has been empirically supported by factor analyses conducted subsequently in this sample.

Conclusion

In individuals meeting DSM-IV criteria for MD in the last year in an epidemiological sample, the individual symptomatic criteria showed strongly varying patterns of relationships with a broad set of clinical relevant validators including demographic characteristics, co-morbidities, characteristics of depressive episodes and personality traits. This pattern is partly reflected by a dichotomy between cognitive and neurovegetative symptoms. Therefore, regarding the characteristics represented by endorsed symptoms, the current DSM-IV definition of MD represents a construct of substantial heterogeneity. These results challenge our understanding of MD as homogenous categorical entity. Hopefully, future research will determine whether a distinction between cognitive and neurovegetative symptoms is clinically relevant to the etiology, course and treatment of MD.

Acknowledgements

This work was supported by the American Psychiatric Association and NIH grants MH-0828 and MH/DA/AA 49492.

Declaration of Interest

None.

References

- Abdi H (2007). Bonferroni and Sidak corrections for multiple comparisons. In *Encyclopedia of Measurement and Statistics* (ed. N. J. Salkind), pp. 103–107. Sage: Thousand Oaks, CA.
- Akiskal HS, McKinney Jr. WT (1975). Overview of recent research in depression. Integration of ten conceptual models into a comprehensive clinical frame. *Archives of General Psychiatry* 32, 285–305.
- Angst J, Dobler-Mikola A (1984). Do the diagnostic criteria determine the sex ratio in depression? *Journal of Affective Disorders* 7, 189–198.
- 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 – Revised*, 3rd edn. American Psychiatric Association: Washington, DC.
- APA (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association: Washington, DC.

- Balazs J, Benazzi F, Rihmer Z, Rihmer A, Akiskal KK, Akiskal HS** (2006). The close link between suicide attempts and mixed (bipolar) depression: implications for suicide prevention. *Journal of Affective Disorders* **91**, 133–138.
- Beck AT, Alford BA** (2008). *Depression: Causes and Treatment*, 2nd edn. University of Pennsylvania Press: Pennsylvania.
- Carragher N, Adamson G, Bunting B, McCann S** (2009). Subtypes of depression in a nationally representative sample. *Journal of Affective Disorders* **113**, 88–99.
- Cassidy WL, Flanagan NB, Spellman M, Cohen ME** (1957). Clinical observations in manic-depressive disease: a quantitative study of one hundred manic-depressive patients and fifty medically sick controls. *Journal of the American Medical Association* **164**, 1535–1546.
- Chen LS, Eaton WW, Gallo JJ, Nestadt G** (2000). Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. *Journal of Affective Disorders* **59**, 1–11.
- Eysenck HJ** (1970). Classification of depressive illnesses. *British Journal of Psychiatry* **117**, 241–250.
- 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.
- Feighner JP, Robins E, Guze SB, Woodruff Jr. RA, Winokur G, Munoz R** (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry* **26**, 57–63.
- Feild H, Armenakis A** (1974). On use of multiple tests of significance in psychological research. *Psychological Reports* **35**, 427–431.
- Frances A, Pincus HA, Widiger TA, Davis WW, First MB** (1990). DSM-IV: work in progress. *American Journal of Psychiatry* **147**, 1439–1448.
- Gillespie RD** (1929). The clinical differentiation of types of depression. *Guy's Hospital Report* **9**, 306–344.
- Hybels CF, Blazer DG, Pieper CF, Landerman LR, Steffens DC** (2009). Profiles of depressive symptoms in older adults diagnosed with major depression: latent cluster analysis. *American Journal of Geriatric Psychiatry* **17**, 387–396.
- Keller MB, Klein DN, Hirschfeld RMA, Kocsis JH, McCullough JP, Miller I, First MB, Holzer CP, Keitner GI, Marin DB, Shea T** (1995). Results of the DSM-IV mood disorders field trial. *American Journal of Psychiatry* **152**, 843–849.
- Keller MC, Neale MC, Kendler KS** (2007). Association of different adverse life events with distinct patterns of depressive symptoms. *American Journal of Psychiatry* **164**, 1521–1529.
- Kendell RE** (1969). Continuum model of depressive illness. *Proceedings of the Royal Society of Medicine – London* **62**, 335–339.
- Kendler KS, Eaves LJ, Walters EE, Neale MC, Heath AC, Kessler RC** (1996). The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Archives of General Psychiatry* **53**, 391–399.
- Kendler KS, Gardner Jr. CO** (1998). Boundaries of major depression: an evaluation of DSM-IV criteria. *American Journal of Psychiatry* **155**, 172–177.
- Kendler KS, Munoz RA, Murphy G** (2009). The development of the Feighner criteria: an historical perspective. *American Journal of Psychiatry*. Published online: 15 December 2009. doi:10.1176/appi.ajp.2009.09081155.
- 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.
- Kendler KS, Zachar P** (2008). The incredible insecurity of psychiatric nosology. In *Philosophical Issues in Psychiatry* (ed. K. S. Kendler and J. Parnas), pp. 368–383. The Johns Hopkins University Press: Baltimore, MD.
- Khan AA, Gardner CO, Prescott CA, Kendler KS** (2002). Gender differences in the symptoms of major depression in opposite-sex dizygotic twin pairs. *American Journal of Psychiatry* **159**, 1427–1429.
- Kiloh LG, Garside RF** (1963). The independence of neurotic depression and endogenous depression. *British Journal of Psychiatry* **109**, 451–463.
- Koenig HG, Cohen HJ, Blazer DG, Krishnan KR, Sibert TE** (1993). Profile of depressive symptoms in younger and older medical inpatients with major depression. *Journal of the American Geriatric Society* **41**, 1169–1176.
- Leventhal AM, Francione WC, Zimmerman M** (2008). Associations between depression subtypes and substance use disorders. *Psychiatry Research* **161**, 43–50.
- Lewis AJ** (1934). Melancholia: a clinical survey of depressive states. *Journal of Mental Science* **80**, 277–378.
- McGlinchey JB, Zimmerman M, Young D, Chelminski I** (2006). Diagnosing major depressive disorder VIII: are some symptoms better than others? *Journal of Nervous and Mental Disease* **194**, 785–790.
- Mapother E** (1926). Discussion on manic-depressive psychosis. *British Medical Journal* **2**, 872–885.
- Merikangas KR, Wicki W, Angst J** (1994). Heterogeneity of depression. Classification of depressive subtypes by longitudinal course. *British Journal of Psychiatry* **164**, 342–348.
- Middeldorp CM, Wray NR, Andrews G, Martin NG, Boomsma DI** (2006). Sex differences in symptoms of depression in unrelated individuals and opposite-sex twin and sibling pairs. *Twin Research and Human Genetics* **9**, 632–636.
- Parker G** (2000). Classifying depression: Should paradigms lost be regained? *American Journal of Psychiatry* **157**, 1195–1203.
- Parker G** (2005). Beyond major depression. *Psychological Medicine* **35**, 467–474.

- Parker G, Wilhelm K, Asghari A** (1998). Depressed mood states and their inter-relationship with clinical depression. *Social Psychiatry and Psychiatric Epidemiology* **33**, 10–15.
- SAS Institute** (2005). SAS OnlineDoc Version 9.1.3. SAS Institute Inc.: Cary, NC.
- Schotte CK, Maes M, Cluydts R, Cosyns P** (1997). Cluster analytic validation of the DSM melancholic depression. The threshold model: integration of quantitative and qualitative distinctions between unipolar depressive subtypes. *Psychiatric Research* **71**, 181–195.
- Spitzer RL, Endicott J, Robins E** (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry* **35**, 773–782.
- Spitzer RL, Fleiss JL** (1974). A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry* **125**, 341–347.
- Stone TT, Burris, BC** (1950). Melancholia; clinical study of 50 selected cases. *Journal of the American Medical Association* **142**, 165–168.
- Sullivan PF, Prescott CA, Kendler KS** (2002). The subtypes of major depression in a twin registry. *Journal of Affective Disorders* **68**, 273–284.
- van Praag HM** (2000). Nosologomania: a disorder of psychiatry. *World Journal of Biological Psychiatry* **1**, 151–158.
- Zimmerman M, McGlinchey JB, Young D, Chelminski I** (2006a). Diagnosing major depressive disorder III: can some symptoms be eliminated from the diagnostic criteria? *Journal of Nervous and Mental Disease* **194**, 313–317.
- Zimmerman M, McGlinchey JB, Young D, Chelminski I** (2006b). Diagnosing major depressive disorder introduction: an examination of the DSM-IV diagnostic criteria. *Journal of Nervous and Mental Disease* **194**, 151–154.