Published in final edited form as:
Bipolar Disord. 2009 May ; 11(3): 289–297. doi:10.1111/j.1399-5618.2009.00681.x.

Differential item functioning of DSM-IV depressive symptoms in individuals with a history of mania versus those without: an item response theory analysis

Lauren M Weinstock, **David Strong**, **Lisa A Uebelacker**, and **Ivan W Miller**
Warren Alpert Medical School, Brown University and Butler Hospital, Providence, RI, USA

Abstract

Objectives—Although major depression is characteristic of both bipolar disorder and major depressive disorder, there is disagreement as to whether there are distinct features of depression that differentiate these two conditions. The primary aim of this study was to use methods based in item response theory to evaluate differences in DSM-IV depression symptom endorsement in an epidemiological sample of individuals with a history of mania (i.e., bipolar depression) versus those without (i.e., unipolar depression).

Methods—Clinical interview data were drawn from a subsample (n = 13,058) of individuals with bipolar or unipolar depression who had participated in the National Epidemiologic Survey on Alcohol and Related Conditions. Using these data, a two-parameter item response model was used to estimate differential item functioning of DSM-IV depressive symptoms between these two groups.

Results—Differences in severity parameter estimates revealed that suicidal ideation and psychomotor disturbance were more likely to be endorsed across most levels of depression severity in bipolar versus unipolar depression. Differences in discrimination parameter estimates revealed that fatigue was significantly less discriminating in bipolar versus unipolar depression.

Conclusions—Equating for level of depression symptom severity, study results revealed that suicidal ideation and psychomotor disturbance are endorsed more frequently in bipolar versus unipolar depression. Study data also suggested that fatigue may be endorsed more frequently in unipolar relative to bipolar samples at moderate (versus low or high) levels of depression symptom severity.

Keywords

bipolar depression; item response theory; unipolar depression

Corresponding author: Lauren M. Weinstock, PhD, Psychosocial Research Program, Warren Alpert Medical School, Brown University and Butler Hospital, 345 Blackstone Boulevard, Providence, RI 02906, USA, Fax: 401-455-6235, lauren_weinstock@brown.edu.

The authors of this paper do not have any commercial associations that might pose a conflict of interest in connection with this manuscript.

Study findings were presented in part at the Seventh International Conference on Bipolar Disorder, Pittsburgh, PA, USA, June 7–9, 2007.

Depression is a key feature of both major depressive disorder (MDD) and bipolar disorder. Although some individuals with bipolar disorder do not experience depression, accumulating evidence suggests that over the long term, the majority of those with bipolar disorder will exhibit depression or depressive symptoms much more frequently than manic symptoms (1, 2). Initial misdiagnosis of MDD is quite common among patients with bipolar disorder (3), which may delay appropriate treatment and, in the case of antidepressant monotherapy, potentially exacerbate mood cycling in vulnerable populations (4). Given this risk for misdiagnosis, it is important to identify whether there are differences in the symptomatic presentation of depression in MDD versus bipolar disorder. If differences exist, understanding them will assist in differential diagnosis and increase the likelihood that a patient receives pharmacological treatment appropriate to his or her illness. Further, an understanding of differences in the phenomenology between bipolar and unipolar depression will be important for creating and adapting psychosocial treatments for bipolar depression.

Although there have been several studies that have attempted to identify differences between depression in MDD versus bipolar disorder, a review of the literature reveals somewhat contradictory and inconclusive findings. Indeed, some (5–9), but not all (10), have reported a greater prevalence of atypical features in bipolar versus unipolar depression. Yet others have reported a greater prevalence of melancholic symptoms in bipolar depression (8, 11). Additional research suggests that unipolar depression may be more strongly characterized by anxiety and somatization (12) and that bipolar depression may be more strongly associated with psychosis (8, 11). Although such related clinical features may, indeed, be informative for differentiating the two syndromes, they are nevertheless not DSM-IV-defined core symptoms of major depression. Thus, such features are limited in informing our understanding of how the core symptoms of depression differentially operate in the two groups. For example, it is possible that a greater prevalence of psychosis in bipolar depression may be better accounted for by a history of mania. Finally, a recent review covering over 30 years of research concluded that unipolar and bipolar depression symptom presentations are more similar than different (12), and others have recently argued that the two cannot be differentiated (13).

This mixed evidence may be related to methodological limitations in the existing literature. Primarily, the large majority of these studies have utilized clinical versus community samples of individuals (12), some of whom had been enrolled in clinical efficacy trials that use narrow inclusion criteria (14, 15). In addition, with few exceptions (11, 15), sample sizes have been generally small, thus rendering estimates unstable. Further, inconsistencies across studies may be accounted for by Type I error, as most studies have performed a large number of statistical comparisons without corresponding alpha level correction (12). Finally, several studies have not controlled for overall symptom severity in their comparisons across groups (5, 6, 11, 12, 14, 16). Thus, it is unclear whether any differential symptom expression is due to true phenomenological differences between bipolar and unipolar depression, or whether such differences are reflective of greater overall depression symptom severity in one group versus another.

To address these limitations, the current study used methods based in item response theory (IRT) (17) to evaluate differences in DSM-IV depression symptom endorsement in a

community sample of individuals with bipolar versus unipolar depression. IRT methodology provides significant improvement over previously used techniques, as it allows one to examine the likelihood that a particular symptom will be endorsed *at a particular level of depression severity*. Thus, differences between bipolar and unipolar depression can be evaluated while simultaneously equating for depression symptom severity. Further extending the literature, we conducted analyses using a large, nationally representative sample of individuals.

Methods

Participants

Participants were drawn from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) (18), a National Institute on Alcohol Abuse and Alcoholism-funded survey of adults in the United States aged 18 years or older. To date, the NESARC represents the largest epidemiological study of psychiatric conditions conducted in the United States. Methods for obtaining the sample have been detailed in other sources (19). In brief, informed consent was obtained from all participants. Only those respondents who reported two weeks of either depressed mood or anhedonia in their lifetime completed the section of the NESARC assessing lifetime occurrence of all DSM-IV symptoms of a major depressive episode (MDE). Of the 43,093 adults surveyed, 1,154 endorsed lifetime depressive symptoms and a lifetime history of manic episodes (i.e., bipolar depression), and 11,904 endorsed lifetime depressive symptoms in the absence of any mania or hypomania (i.e., unipolar depression). The present analysis consisted of only those individuals (n = 13,058; 30% of the total sample). For participants with bipolar and unipolar depression, respectively, average age was 39.4 (SD = 14.8) and 47.1 (SD = 17.2[1]. Among participants with bipolar depression, 63% (n = 727) were female[2], 78% (n = 898) were Caucasian, and 83% (n = 958) were of non-Hispanic ethnicity. Among participants with unipolar depression, 66% (n = 7,857) were female, 80% (n = 9,544) were Caucasian, and 84% (n = 9,999) were of non-Hispanic ethnicity.

Assessments

The Alcohol Use Disorders and Associated Disabilities Interview Schedule–DSM-IV version (AUDADIS-IV) (20) was used to assess MDE symptoms and manic episode criteria. Experienced interviewers received extensive training in this fully structured interview and used computer-assisted software in order to decrease error in measurement (19). Developers of the AUDADIS-IV also made considerable efforts to ensure that questions were comprehensible for laypersons (21). Extensive data concerning the psychometric performance of the AUDADIS-IV have been reported elsewhere (21). NESARC estimates of lifetime and 12-month prevalence of MDD were 13.2% and 5.3%, respectively. NESARC

[1]Although it is possible that the younger age of the bipolar sample might have influenced study results, it is noteworthy that this group reported overall *greater* lifetime symptomatology than those with unipolar depression (see Table 1). Thus, it is not likely that IRT analyses were confounded by age.

[2]Of the 41,682 individuals who participated in the NESARC, 57% (n = 23,743) were women. When overall prevalence of bipolar I disorder was evaluated, including those with unipolar mania, prevalence of lifetime bipolar I disorder was 3.4% and 3.1% for women and men, respectively. A sample-weighted comparison revealed no gender difference in the prevalence of lifetime bipolar I disorder ($\chi^2 = 2.43$, df = 1, p = 0.21), which is consistent with prior reports.

estimates of lifetime and 12-month prevalence of bipolar I disorder were 3.3% and 2.0%, respectively. These estimates are generally comparable to those found in other recent epidemiological surveys (22), although it should be noted that the prevalence rates for bipolar I disorder in the NESARC "slightly exceeded the upper end of the range" of previously reported estimates (23, p. 1211). Test-retest reliability for the MDD (24) and bipolar I disorder (23) diagnoses was good in this sample.

For purpose of analysis, bipolar depression was characterized by endorsement of lifetime threshold-level manic episode criteria and endorsement of lifetime MDE symptoms. Unipolar depression was characterized by endorsement of lifetime MDE symptoms in the absence of any manic or hypomanic episodes. Analyses focused on the seven MDE symptoms assessed by the AUDADIS-IV once depressed mood and/or anhedonia were endorsed: appetite/weight disturbance, sleep disturbance, psychomotor disturbance, fatigue, worthlessness/guilt, concentration difficulty, suicidal ideation/attempt.

Statistical analyses

Overview—In order to estimate differential item functioning (DIF) in bipolar versus unipolar depression, a two-parameter item response model was used. Item response modeling allows us to examine the likelihood that a particular symptom will be reported at a particular level of depression severity (i.e., the latent trait) in different groups. As this method ensures that individual characteristics do not affect interpretation of total symptom counts, equal comparisons can be made across groups. The estimation of DIF involved comparing a series of analyses that isolate and compare each item parameter across each identified group (25, 26). If the symptoms function similarly across groups, then the parameters that describe the symptoms of depression will be estimated similarly in different samples.

IRT model—Parametric models begin with a specific model of how the relationship between the probability of an item response and an individual's level of the underlying trait should look (e.g., the item response function), and then model the estimated parameters that describe the relationship. For purposes of this analysis, a two-parameter item response model was selected. This model estimates: (i) a severity parameter to describe the point on the latent continuum where a symptom becomes likely to be observed (e.g., > 50%); and (ii) a discrimination parameter to describe how rapidly the probability of observing the symptom changes across increasing levels of the latent continuum (e.g., slope of the item response function). It is important to acknowledge that, in the current study, interpretation of potential severity DIF is most relevant to study aims. Although we used a two-parameter model because it provided a better fit to the data than a one-parameter model, we were most interested in DIF that occurred in the severity parameter, as that is reflective of the likelihood that a given symptom will occur at a given severity level. However, the discrimination parameter is important in that it can be used to verify that a given symptom is a good indicator of the underlying latent dimension.

Unidimensionality assumption—The primary assumption of item response models is that responses to symptom queries are a function of individual variation along a single

underlying dimension, which we tested using confirmatory common factor analysis of tetrachoric correlations. This assumption is meaningful for both theoretical and statistical reasons. Theoretically, the DSM-IV stipulates that symptoms are summed to determine the presence or absence of a depression diagnosis. In so doing, DSM-IV assumes that responses are linked to a single construct of depression severity (27). Statistically, information regarding symptom functioning may be biased if a unidimensional item response model is applied to multidimensional data.

Differential item functioning—We employed Version 2.0 of IRTLRDIF (D. Thissen, IRTLRDIF v. 2.0b: Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio Tests for Differential Item Functioning, 2001, unpublished manuscript) to complete DIF analyses. IRTLRDIF automatically accommodates group differences with respect to the latent trait. IRTLRDIF sets the scale of item parameters using the population distribution for the reference group. With the reference group mean set at zero and standard deviation set at one, the estimated focal group mean reflects a standardized difference from the reference group and the standard deviation reflects the ratio of the focal and reference group standard deviations (D. Thissen, 2001, unpublished manuscript).

Following Thissen et al. (28), we used a likelihood-ratio test statistic to provide a significance test for the null hypothesis that the item parameters do not differ between the identified groups (i.e., bipolar and unipolar depression). Analyses proceeded by initially constraining both discrimination and severity estimates to be equal for the two subgroups across all seven symptoms (Model A). For each of the seven symptoms, a model was then fit that constrains all of the remaining symptoms' discrimination and severity estimates to be equal, but allows the estimates for one symptom to differ across the two groups (Model B). The difference in the log-likelihoods (*ll*) of Model A and Model B [$G^2 = -2(ll_{\text{Model A}} - ll_{\text{Model B}})$] provides an omnibus test (df = 2) of whether there is DIF for the discrimination and/or severity estimate for this symptom. If significant, follow-up tests can be conducted to identify whether DIF is present in discrimination or severity estimates by further constraining models.

Given that we conducted DIF analyses across multiple symptoms, it is important to account for risk of Type I error. Although Bonferroni correction has typically been used to do so, this strategy can be conservative and may result in reduced power to detect differences. As an alternative, we employed the Benjamini-Hochberg procedure (29), using methods described elsewhere (30). For all DIF analyses, we set alpha at 0.05. We used the Benjamini-Hochberg procedure to adjust p-values for all 1*df* tests.

Given the large sample size employed in the current study, relatively small differences between groups may emerge as statistically significant. A priori, we decided that a difference ≥ 0.25 in item severity would represent a clinically meaningful difference. As described by Steinberg and Thissen (31), a difference of 0.25 can be interpreted as one quarter of a "standard unit difference between the values of the [underlying] trait necessary to have a 50–50 chance of responding positively in one group compared to another" (31, pp. 405–406). This may be considered to be a small effect size (32). For example, a DIF of 0.25

for a given item severity would mean that, depending on the values of the discrimination parameters as well as how close the actual group severity parameters are to 0, differences in group proportions responding affirmatively to a given item could range from 2% to 8% (for discrimination parameters ranging from 0.50 to 2.00) (31).

With respect to discrimination parameters, Steinberg and Thissen (31) suggest that the best way to determine whether a statistically significant discrimination parameter is also clinically significant is by visual inspection of the respective item response functions. This recommendation is based upon the argument that interpretation of discrimination DIF outside the context of overall severity (as represented by the *b* parameter) provides limited information regarding the clinical utility of the effect. In contrast, a "graphic display is the most easily comprehensible description of the effect size," (31, p. 411) because it allows for the interpretation of discrimination DIF relative to severity. Interested readers are referred to Steinberg and Thissen (31) for an expanded discussion of effect size estimation in two-parameter IRT models.

Results

Unidimensionality assumption

We conducted confirmatory factor analyses in order to test the assumption of the unidimensionality of depression symptoms in the unipolar and bipolar subgroups. Fit statistics for the unipolar group [$X^2 = 543.95$; comparative fit index (CFI) = 0.977; Tucker-Lewis Index (TLI) = 0.978; root mean square error of approximation (RMSEA) = 0.059] and bipolar group ($X^2 = 62.11$; CFI = 0.945; TLI = 0.953; RMSEA = 0.057) indicated a reasonable fit to the data. We determined that these fit statistics were sufficient to proceed to fitting IRT models.

Differential item functioning

Table 1 lists frequency of endorsement of each of the DSM-IV major depression symptoms for each participant group, as well as corresponding severity and discrimination parameter estimates. As reflected in the generally greater frequency of symptom endorsement among those with bipolar depression (see Table 1), there was an overall depression severity difference between the two groups, with the bipolar group having a mean depression severity that was 1.17 standard deviation units higher than the unipolar group. It is important to note that this overall difference is accounted for in the DIF analysis by equating for depression severity on the latent dimension.

Table 2 lists group differences in severity and discrimination parameter estimates for each symptom. As evidenced in Table 2, two items exceeded study criteria for statistically and clinically significant severity DIF. One item exceeded study criteria for both statistically and clinically significant discrimination DIF. To aid in interpretation of findings, Figure 1 represents a pictorial representation of DIF for all seven of the depression symptoms evaluated. Differences in severity parameter estimates for bipolar and unipolar depression revealed that psychomotor disturbance ($b_{bipolar} = -0.11$, $b_{unipolar} = 0.17$, $b_{difference} = -0.28$) and suicidal ideation ($b_{bipolar} = 0.01$, $b_{unipolar} = 0.48$, $b_{difference} = -0.47$) were endorsed at

lower levels of depression severity in bipolar versus unipolar depression. The differences in severity parameter estimates were statistically significant for both psychomotor disturbance (G^2 = 5.4, df = 1, p < 0.001) and suicidal ideation (G^2 = 27.3, df = 1, p < 0.001).

Data revealed significant differences in discrimination parameter estimates for fatigue ($a_{bipolar}$ = 1.34, $a_{unipolar}$ = 1.85, $a_{difference}$ = −0.51; G^2 = 8.5, df = 1, p < 0.01) and psychomotor disturbance ($a_{bipolar}$ = 1.41, $a_{unipolar}$ = 1.81, $a_{difference}$ = −0.40; G^2 = 5.1, df = 1, p < 0.01). However, visual inspection of the item response functions revealed that psychomotor disturbance was endorsed consistently at lower levels of severity in bipolar relative to unipolar depression (i.e., the item response curves cross only at the highest levels of depression severity). It is only at the most severe levels of depression that non-uniform DIF could potentially impact interpretation of the model, but at that point on the distribution, nearly 100% of the current study sample endorsed psychomotor disturbance (see Fig. 1). Thus, the effect of discrimination DIF on this item is seemingly minimal, and therefore appears to be of limited clinical significance.

Discussion

In an effort to better understand differences between unipolar and bipolar depression, the aim of the current study was to evaluate differential functioning of DSM-IV depression symptoms in these groups using methods based on item response theory. Strengths of this study include the use of an IRT-based methodology, which accounts for the potential confounding effect of depression severity in evaluating group differences, and the use of a large, representative community sample of individuals. Additional strengths of this study include the use of the Benjamini-Hochberg procedure to correct for multiple comparisons, and the establishment of an a-priori effect size threshold for interpretation so that an emphasis was placed on differences that were both statistically significant and clinically meaningful.

Consistent with the conclusions reached in several large-scale, comprehensive reviews (12, 33, 34), current study results revealed relatively few differences in symptom expression between unipolar and bipolar depression. Among the differences that did emerge, data suggested that suicidal ideation and psychomotor disturbance may be more common in bipolar depression, and that fatigue may be more discriminating in unipolar depression. Ranging in size from small to medium in magnitude, these effects suggest that there may be subtle differences in depression symptom presentation that may nevertheless be informative in distinguishing between unipolar and bipolar depression. The advantage of the size and nature of our sample, as well as our statistical methods, is that we were able to detect subtle but real differences, which are discussed in greater detail below.

In the current study, individuals with bipolar depression were significantly more likely to endorse suicidal ideation at lower levels of depression severity. A visual inspection of the item response functions suggests that, *at an average level of depression severity*, approximately 50% of individuals with bipolar depression will endorse suicidal ideation, in comparison to approximately 30% of individuals with unipolar depression (see Fig. 1). These data are consistent with prior findings that suicidal ideation and rates of suicide

attempt are higher in bipolar disorder versus MDD (15, 35–37), and may be the highest of any Axis I disorder (36). Given that the large majority of suicide attempts in bipolar disorder occur during a major depressive episode (38), this finding is also consistent with established data linking depression in particular to suicide in bipolar samples. As the DIF analysis equates for depression severity between groups, current study data further extend the literature by suggesting that differences in suicidal ideation between bipolar and unipolar depression were not merely due to greater depressive symptom severity in bipolar disorder. Rather, there may exist other meaningful differences, such as mixed depressive states in bipolar disorder (39), greater hopelessness secondary to having a "severe mental illness," or more frequent hospitalizations that may account for differences in suicide between bipolar and unipolar depression. Future research that further explores such differences may improve our ability to successfully prevent suicide in the context of bipolar disorder.

Additional study findings revealed that bipolar depression was more likely to be characterized by psychomotor disturbance. Inasmuch as individuals with bipolar disorder may be particularly vulnerable to atypical (6–9) or irritable (40) depression, and that such individuals also have a history of mania, it may not be surprising that they present with psychomotor difficulties at lower levels of depression severity. Although prior research has indicated a specific link between psychomotor retardation and bipolar depression (9, 16, 35), given the study data, it is unclear whether those with bipolar depression exhibit greater levels of retardation, agitation, or both. Although data were collected for each individually in the NESARC, one explicit assumption of IRT methodology is that symptoms be locally independent (i.e., they must not be correlated for reasons other than measurement of the latent construct) (17). Because one could reliably predict the absence of one from the presence of the other, irrespective of depression severity, we can assume that psychomotor agitation and retardation are locally dependent, and thus not appropriate for evaluation separately in an IRT analysis. Therefore, the symptoms were evaluated together as a compound item that directly parallels the larger DSM-IV criterion of psychomotor disturbance that is used in the assignment of MDE diagnosis. Thus, future research will be necessary to further clarify the nature of psychomotor disturbance in bipolar, relative to unipolar, depression.

Results from analysis of discrimination DIF revealed that fatigue was significantly less discriminating in bipolar versus unipolar depression. Although these data suggest that fatigue may be less useful as an indicator of depression severity in bipolar samples, a visual inspection of the item response functions nonetheless reveals that fatigue remains a reasonable discriminator of depression severity. Perhaps more informative is the interpretation of discrimination DIF relative to severity DIF. Indeed, the non-uniform DIF for fatigue may at least partially explain why statistically significant severity DIF failed to meet our threshold for clinical significance. In particular, at both low and high depression severity, unipolar and bipolar samples endorsed fatigue at similar rates. However, *at moderate levels of depression severity*, individuals with unipolar depression endorsed fatigue more frequently (see Fig. 1). Such data illustrate why it is critical to account for overall depression severity when evaluating differential symptom expression between clinical groups.

When interpreting the findings above, it is important to acknowledge current study limitations. First, in order to be included in the data analysis, individuals must have endorsed either depressed mood or anhedonia. In the NESARC, if a respondent did not endorse either, the remaining DSM-IV depression symptoms were not asked. Thus, given that a majority of the current study sample endorsed one or both of these symptoms, it would not be terribly meaningful, from a statistical perspective, to conduct DIF analyses on these items. Furthermore, there may be other clinical features that differentiate unipolar and bipolar depression that are otherwise not DSM-IV symptoms of depression (e.g., anxiety) or that were not assessed in the NESARC (e.g., mood reactivity). Thus, it is important to emphasize that study results cannot be generalized beyond the core symptoms of major depression, and we cannot generalize results to form conclusions about depression subtypes, such as melancholia or atypical depression. To the extent that we did have data on atypical symptoms (i.e., hypersomnia and hyperphagia), we could not evaluate these symptoms separately from their counterparts (i.e., insomnia and loss of appetite) due to assumptions of local independence, as described above. Additionally, there may be other clinical course characteristics (e.g., age of onset, rates of recurrence or hospitalization, or medication regimen) that might potentially influence symptom profiles. Finally, current study results cannot be generalized to form conclusions about bipolar II disorder. As bipolar II depression may be particularly difficult to discern from unipolar depression (41), further research in this area is warranted.

In sum, equating for depression symptom severity, current study data revealed relatively few differences in symptom expression in bipolar versus unipolar depression. Among the symptoms that did appear to differentiate these groups, suicidal ideation and psychomotor disturbance were endorsed more frequently in bipolar depression. Notably, the effect size for the difference in suicidal ideation endorsement was the largest in the current study, which further highlights the importance of monitoring suicidal ideation in patients with depression, especially those with a history of mania. Further, significant non-uniform DIF suggested that fatigue was endorsed more frequently in unipolar relative to bipolar samples at moderate, but not at low or high levels of depression symptom severity. Study results may be useful in aiding diagnostic decisions among clinicians, and may highlight subtle phenomenological differences between bipolar and unipolar depression that can be used to guide future research.

Acknowledgments

The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) was supported by grants R01AA08159 and K05AA00161 from the National Institute on Alcohol Abuse and Alcoholism. Preparation of this manuscript was supported by grant K23MH079907 from the National Institute of Mental Health, awarded to LMW.

References

1. Judd LL, Akiskal HS, Schettler PJ, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry. 2002; 59:530–537. [PubMed: 12044195]
2. Post RM, Denicoff KD, Leverich GS, et al. Morbidity in 258 bipolar outpatients followed for 1 year with daily prospective ratings on the NIMH life chart method. J Clin Psychiatry. 2003; 64:680–690. [PubMed: 12823083]

3. Hirschfeld RM, Lewis L, Vornik LA. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. J Clin Psychiatry. 2003; 64:161–174. [PubMed: 12633125]
4. Leverich GS, Altshuler LL, Frye MA, et al. Risk of switch in mood polarity to hypomania or mania in patients with bipolar depression during acute and continuation trials of venlafaxine, sertraline, and bupropion as adjuncts to mood stabilizers. Am J Psychiatry. 2006; 163:232–239. [PubMed: 16449476]
5. Abrams R, Taylor MA. A comparison of unipolar and bipolar depressive illness. Am J Psychiatry. 1980; 137:1084–1087. [PubMed: 6107049]
6. Benazzi F. Atypical depression in private practice depressed outpatients: a 203-case study. Compr Psychiatry. 1999; 40:80–83. [PubMed: 9924882]
7. Bowden CL. A different depression: clinical distinctions between bipolar and unipolar depression. J Affect Disord. 2005; 84:117–125. [PubMed: 15708408]
8. Mitchell PB, Wilhelm K, Parker G, Austin MP, Rutgers P, Malhi GS. The clinical features of bipolar depression: a comparison with matched major depressive disorder patients. J Clin Psychiatry. 2001; 62:212–216. [PubMed: 11305713]
9. Potter WZ. Bipolar depression: specific treatments. J Clin Psychiatry. 1998; 59(Suppl):30–36. [PubMed: 9840196]
10. Robertson HA, Lam RW, Stewart JN, Yatham LN, Tam EM, Zis AP. Atypical depressive symptoms and clusters in unipolar and bipolar depression. Acta Psychiatr Scand. 1996; 94:421–427. [PubMed: 9020993]
11. Parker G, Roy K, Wilhelm K, Mitchell P, Hadzi-Pavlovic D. The nature of bipolar depression: implications for the definition of melancholia. J Affect Disord. 2000; 59:217–224. [PubMed: 10854638]
12. Cuellar AK, Johnson SL, Winters R. Distinctions between bipolar and unipolar depression. Clin Psychol Rev. 2005; 25:307–339. [PubMed: 15792852]
13. Joffe RT, Young LT, MacQueen GM. A two-illness model of bipolar disorder 1. Bipolar Disord. 1999; 1:25–30. [PubMed: 11256651]
14. Katz MM, Robins E, Croughan J, Secunda S, Swann A. Behavioural measurement and drug response characteristics of unipolar and bipolar depression. Psychol Med. 1982; 12:25–36. [PubMed: 7079431]
15. Perlis RH, Brown E, Baker RW, Nierenberg AA. Clinical features of bipolar depression versus major depressive disorder in large multicenter trials. Am J Psychiatry. 2006; 163:225–231. [PubMed: 16449475]
16. Brockington IF, Altman E, Hillier V, Meltzer HY, Nand S. The clinical picture of bipolar affective disorder in its depressed phase. A report from London and Chicago. Br J Psychiatry. 1982; 141:558–562. [PubMed: 7159802]
17. Lord, FM. Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Erlbaum; 1980.
18. Grant, BF.; Kaplan, K.; Shepard, J.; Moore, T. Source and Accuracy Statement for Wave 1 of the 2001–2002 National Epidemiologic Survey or Alcohol and Related Conditions. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2003. Available at: http://www.niaaa.nih.gov [Accessed on October 1, 2007]
19. Grant BF, Stinson FS, Dawson DA, Chou SP, Ruan WJ, Pickering RP. Co-occurrence of 12-month alcohol and drug use disorders and personality disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry. 2004; 61:361–368. [PubMed: 15066894]
20. Grant, BF.; Dawson, DA.; Hasin, DS. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2001. Available at: http://www.niaaa.nih.gov [Accessed on January 15, 2007]
21. Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. Drug Alcohol Depend. 2003; 71:7–16. [PubMed: 12821201]

22. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005; 62:617–627. [PubMed: 15939839]
23. Grant BF, Stinson FS, Hasin DS, et al. Prevalence, correlates, and comorbidity of bipolar I disorder and axis I and II disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry. 2005; 66:1205–1215. [PubMed: 16259532]
24. Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. Arch Gen Psychiatry. 2005; 62:1097–1106. [PubMed: 16203955]
25. Camilli, G.; Shepard, LA. Methods for Identifying Biased Test Items. Thousand Oaks, CA: Sage Publications; 1994.
26. Holland, PW.; Wainer, H. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993.
27. Thurstone LL. Attitudes can be measured. Amer J Sociology. 1928; 33:529–554.
28. Thissen, D.; Steinberg, L.; Wainer, H. Detection of differential item functioning using the parameters of item response models. In: Holland, PW.; Wainer, H., editors. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993. p. 67-113.
29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Royal Statistical Society, Series B. 1995; 57:289–300.
30. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. Journal of Educational and Behavioral Statistics. 2002; 27:77–83.
31. Steinberg L, Thissen D. Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. Psychological Methods. 2006; 11:402–415. [PubMed: 17154754]
32. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. 2nd Ed.. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
33. Cassano GB, Akiskal HS, Musetti L, Perugi G, Soriani A, Mignani V. Psychopathology, temperament, and past course in primary major depressions. 2. Toward a redefinition of bipolarity with a new semistructured interview for depression. Psychopathology. 1989; 22:278–288. [PubMed: 2602525]
34. Swann, AC. Is bipolar depression a specific biological entity?. In: Young, LT.; Joffe, RT., editors. Bipolar Disorder: Biological Models and Their Clinical Application. New York: Dekker; 1994. p. 225-286.
35. Mitchell PB, Malhi GS. Bipolar depression: phenomenological overview and clinical characteristics. Bipolar Disord. 2004; 6:530–539. [PubMed: 15541069]
36. Oquendo MA, Mann JJ. Suicidal behavior in bipolar disorder: state of the biological, clinical, and interventions evidence. Bipolar Disord. 2006; 8:523–525.
37. Mitchell PB, Slade T, Andrews G. Twelve-month prevalence and disability of DSM-IV bipolar disorder in an Australian general population survey. Psychol Med. 2004; 34:777–785. [PubMed: 15500298]
38. Tondo L, Isacsson G, Baldessarini R. Suicidal behaviour in bipolar disorder: risk and prevention. CNS Drugs. 2003; 17:491–511. [PubMed: 12751919]
39. Balazs J, Benazzi F, Rihmer Z, Rihmer A, Akiskal KK, Akiskal HS. The close link between suicide attempts and mixed (bipolar) depression: implications for suicide prevention. J Affect Disord. 2006; 91:133–138. [PubMed: 16458364]
40. Deckersbach T, Perlis RH, Frankle WG, et al. Presence of irritability during depressive episodes in bipolar disorder. CNS Spectr. 2004; 9:227–231. [PubMed: 14999163]
41. Benazzi F. A continuity between bipolar II depression and major depressive disorder? Prog Neuropsychopharmacol Biol Psychiatry. 2006; 30:1043–1050. [PubMed: 16682104]

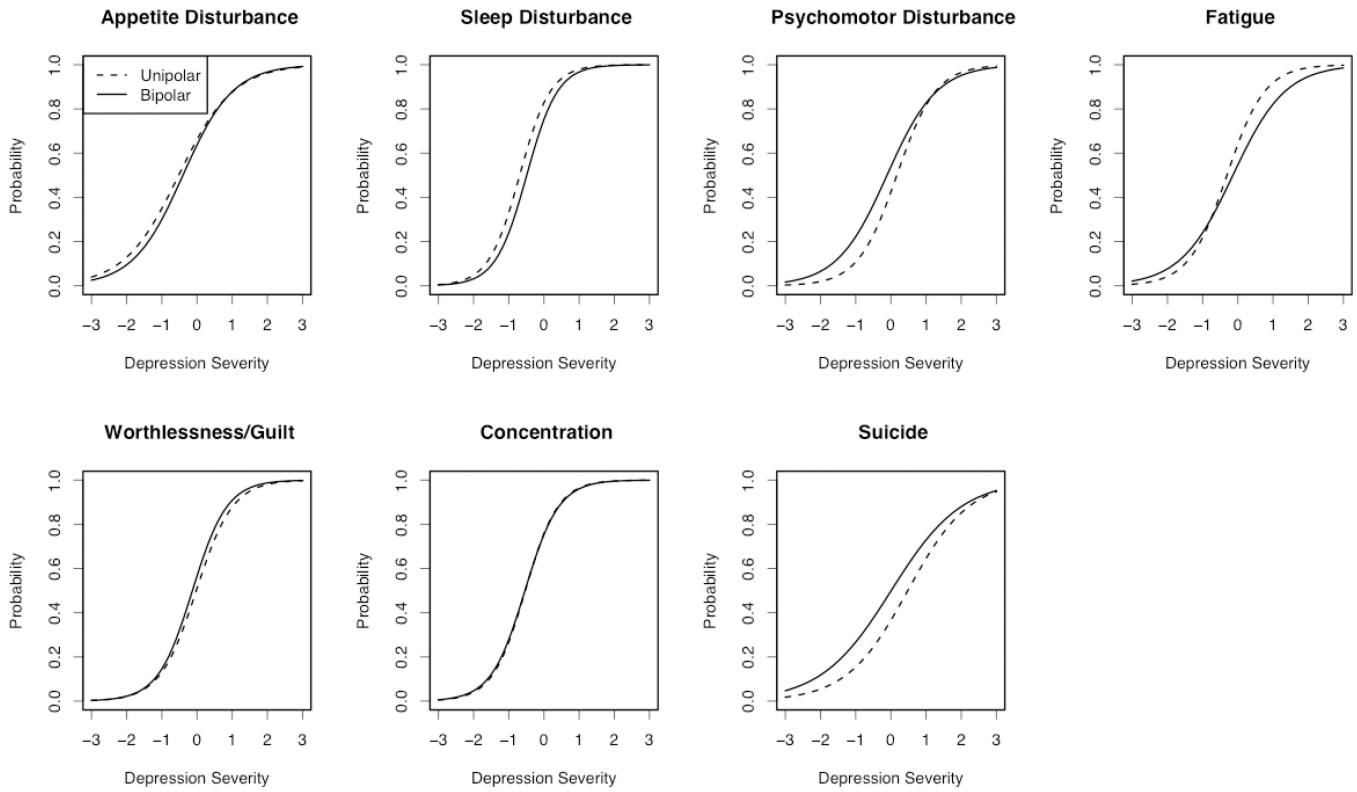

Fig. 1.

Differential item functioning between unipolar and bipolar depression for DSM-IV major depressive episode symptoms

Table 1

Initial symptom parameters for study participants (n = 13,058)

	Bipolar depression (n = 1,154)			Unipolar depression (n = 11,904)		
DSM-IV major depressive episode symptoms	**%**	***a***	***b***	**%**	***a***	***b***
Depressed mood	92.8	–	–	93.1	–	–
Anhedonia	91.6	–	–	74.6	–	–
Appetite/weight disturbance	84.4	1.41	−0.40	62.6	1.29	−0.52
Decreased appetite/weight loss	63.4	–	–	45.0	–	–
Increased appetite/weight gain	44.0	–	–	25.4	–	–
Sleep disturbance	91.2	2.27	−0.49	71.2	2.28	−0.69
I nsomnia	79.6	–	–	58.9	–	–
Hypersomnia	52.0	–	–	31.1	–	–
Psychomotor disturbance	79.1	1.41	−0.11	45.2	1.81	0.17
Psychomotor agitation	68.7	–	–	34.4	–	–
Psychomotor retardation	47.3	–	–	24.5	–	–
Fatigue	80.4	1.34	−0.14	59.3	1.85	−0.31
Worthlessness/guilt	84.0	2.03	−0.13	51.0	1.93	−0.02
Concentration difficulty	90.7	2.07	−0.54	66.4	2.15	−0.53
Suicidal ideation	71.9	1.00	0.01	39.4	1.15	0.48

a discrimination parameter estimate;

b severity parameter estimate. Frequencies for component parts of appetite, sleep, and psychomotor disturbance are presented for descriptive purposes only. Due to assumptions of local independence required for item response theory analysis, these component items were not evaluated separately.

Table 2

Differential item functioning (DIF)

	G^2 (df = 2)	Severity DIF	Discrimination DIF
Appetite/weight disturbance	0.9	0.12	0.12
Sleep disturbance	10.0[a]	0.20	−0.01
Psychomotor disturbance	10.4[a]	**−0.28**[b]	−0.40[a]
Fatigue	50.3[b]	0.17	**−0.51**[a]
Worthlessness/guilt	6.1[c]	−0.11	0.10
Concentration difficulty	0.2	−0.01	−0.08
Suicidal ideation	28.4[b]	**−0.47**[b]	−0.15

Differences between the groups on either parameter were evaluated using 1*df* tests. The p-values for the 1*df* tests were adjusted using the Benjamini-Hochberg procedure. Parameters that (i) represent a statistically significant difference between groups, and (ii) meet our criteria for clinical significance are **bolded**.

[a] $p < 0.01$;

[b] $p < 0.001$;

[c] $p < 0.05$.