



# HHS Public Access

Author manuscript

*J Affect Disord.* Author manuscript; available in PMC 2017 August 01.

Published in final edited form as:

*J Affect Disord.* 2016 August ; 200: 67–73. doi:10.1016/j.jad.2016.04.034.

## Differential endorsement of suicidal ideation and attempt in bipolar versus unipolar depression: a testlet response theory analysis

Lauren M. Weinstock<sup>a,b,\*</sup>, David Strong<sup>c</sup>, Lisa A. Uebelacker<sup>a,b</sup>, and Ivan W. Miller<sup>a,b</sup>

<sup>a</sup> Alpert Medical School of Brown University, Department of Psychiatry & Human Behavior, Providence, RI, USA

<sup>b</sup> Butler Hospital, Psychosocial Research Program, Providence, RI, USA

<sup>c</sup> University of California, San Diego, Department of Family Medicine and Public Health, San Diego, CA, USA

### Abstract

**Background**—Published data concerning differences in suicide risk across the mood disorders spectrum remain mixed. The current study used testlet response theory methods to evaluate differences in the endorsement of suicidal ideation and attempt in an epidemiological sample of individuals with bipolar and unipolar depression.

**Method**—Participants with lifetime history of bipolar I ( $n=1154$ ), bipolar II ( $n=494$ ), and unipolar ( $n=5695$ ) depression were drawn from the National Epidemiologic Survey on Alcohol and Related Conditions, which included 4 structured queries concerning suicidal ideation/attempt. We estimated differential item functioning between groups with a 2-pl parametric item response model.

**Results**—Endorsement of suicide items increased as a function of underlying depression severity. Equating for severity, endorsement of suicidal ideation and attempt was generally more frequent in bipolar versus unipolar depression, and in bipolar I versus bipolar II depression. Yet findings were not consistent across all suicide items, and differences were small in magnitude.

**Limitations**—The NESARC relied upon lifetime endorsement of suicide items, and suicide risk was only evaluated within the context of a major depressive episode. Thus, this study could not evaluate endorsement of suicide items within the context of (hypo)manic or mixed states.

**Conclusion**—Although there were some group differences, patterns of item endorsement were more similar than different. These data support a transdiagnostic model of suicide that emphasizes underlying depression severity over mood disorder class.

### Keywords

Suicide; Bipolar disorder; Major depressive disorder; Testlet response theory

---

\* Corresponding author at: Alpert Medical School of Brown University and Butler Hospital, 345 Blackstone Boulevard, Providence, RI 02906, USA. ; Email: Lauren\_Weinstock@brown.edu (L.M. Weinstock).

## 1. Introduction

Apart from suicide attempt history, research suggests that the strongest predictor of suicide is the presence of a major depressive episode (MDE; Rihmer and Kiss, 2002; Tondo et al., 2003). Data compiled from national and international psychological autopsy studies reveal that 90% of suicide deaths are associated with one or more major psychiatric diagnoses, with MDE (59–87%) representing the most frequent principal diagnosis (Rihmer and Kiss, 2002). Yet MDEs are expressed across several different psychiatric disorders, most notably major depressive disorder (MDD) and bipolar disorder (BD), and there remains disagreement as to whether there are differences in the expression of suicidal ideation and behaviors between those with MDD and BD (Holma et al., 2014; Weinstock et al., 2010b). Understanding such differences, should they exist, is critical for identifying those individuals who are at highest risk in an otherwise high risk group for suicide-related outcomes.

In their review of the literature, Rihmer and Kiss (2002, p. 21) concluded that, “bipolar patients in general, and bipolar II subjects in particular, carry the highest risk of suicide.” Yet the published data do not fully support this assertion. Some have reported greater suicide risk in BD relative to MDD (Bottlender et al., 2000; Chen and Dilsaver, 1996; Raja and Azzoni, 2004; Tondo et al., 2007), whereas others have reported the opposite effect, with greater suicide risk in MDD (Angst et al., 2002; Lester, 1993). Still others have failed to find differences between BD and MDD on levels of suicidal ideation (Bottlender et al., 2000; Zalsman et al., 2006), number of suicide attempts (Fiedorowicz et al., 2009; Holma et al., 2014; Zalsman et al., 2006), or intent to die (Zalsman et al., 2006). Within BD, there is some evidence for the assertion that suicide risk is higher in bipolar II disorder (BDII) versus bipolar I disorder (BDI), with reports of higher lifetime history of suicide attempts (Moreno and Andrade, 2005; Tondo et al., 1999) and suicide deaths (Tondo et al., 2007) in BDII. Yet there is also evidence that lifetime history of suicide attempt is higher in BDI relative to BDII (Moreno et al., 2012; Tondo et al., 2007; Weinstock et al., 2010a), and several additional studies have failed to find differences between BD subtypes on suicide risk indices such as family history of suicide, suicidal ideation, or suicide attempts (Cassano et al., 1992; Coryell et al., 1987; Endicott et al., 1985; Vieta et al., 1997).

These mixed empirical findings may be attributable to a number of limitations in the extant literature. First, several studies have relied on combined samples of BDI and BDII for purposes of comparison against MDD (Angst et al., 2002; Bottlender et al., 2000; Fiedorowicz et al., 2009; Raja and Azzoni, 2004; Zalsman et al., 2006), which may lead to inconsistencies across studies, especially if there are differences in suicide risk between bipolar subtypes. Second, with some exceptions (Moreno et al., 2012; Moreno and Andrade, 2005; Weinstock et al., 2010a), there has been a heavy reliance on clinical versus community samples, which may limit variability of and potentially bias suicide outcomes. Finally, and perhaps most importantly, the majority of studies have failed to account for potential group differences in underlying depression severity. As such, it remains unclear whether any reported differences in suicidal risk across BDI, BDII, and MDD were due to true phenomenological differences between these groups, or whether such differences were instead reflective of greater depression severity in any one group versus another (Weinstock et al., 2010b, 2013). This limitation is especially critical in light of the heterogeneity in

sample selection (i.e., inpatient versus outpatient, clinical versus community) that may impact clinical severity within and between published studies. Indeed, when data from two separate clinical cohorts of psychiatric patients with BD (including BDI and BDII) and MDD were recently evaluated together, there were no diagnostic differences in prospective reports of suicidal ideation or attempts once depression severity was accounted for in the statistical models (Holma et al., 2014).

We have previously argued (Weinstock et al., 2009, 2010a, 2010b, 2013) that methods based in item response theory (IRT; Lord, 1980) may be particularly useful in overcoming some of the challenges that have plagued the existing literature on differential suicide risk across the mood disorders. The primary advantage of an IRT-based approach over other statistical methodologies is that it allows one to examine the likelihood that a particular item (e.g., suicidal ideation) will be endorsed at a particular level of depression severity. Thus, differences in symptom endorsement between groups can be evaluated while simultaneously equating for underlying depression severity. In an application of this concept, we relied upon methods based in IRT to evaluate differences in the endorsement of depressive symptoms, including suicidal ideation, across bipolar I, bipolar II, and unipolar depression in a large, epidemiological sample (Weinstock et al., 2009, 2010a). Contrary to the belief that suicide risk might be highest among BDII, results from these analyses revealed that individuals with bipolar I depression were more likely to endorse suicidal thoughts and behaviors than those with bipolar II or unipolar depression (Weinstock et al., 2010a). The data further revealed no differences in the endorsement of suicidal ideation between bipolar II and unipolar depression (Weinstock et al., 2010a). However, due to statistical assumptions of traditional IRT, which cannot accommodate items that are locally dependent (i.e., correlated for reasons other than measurement of the underlying trait) (Lord, 1980), one limitation of this research was that it relied on a composite measure of suicidality that collapsed across reports of suicidal thoughts, intent, and attempts. As such, we were unable to evaluate potential differences between groups across specific indices of suicide risk, which may have limited an ability to detect more nuanced differences between diagnostic groups.

In an effort to further advance this line of inquiry, the primary aim of the current study was to evaluate differential endorsement of suicide risk indices (e.g., thoughts of death, thoughts of suicide, suicide attempt) across BDI, BDII, and MDD, capitalizing on recent methodological advances in IRT that allow for the accommodation of local dependence among related items. Using these methods, based in testlet response theory (TRT; Baldwin et al., 2007; Wainer et al., 2007), we estimated differential item functioning (DIF; Thissen et al., 1993) across diagnostic groups drawn from a large, community-based sample of individuals.

## 2. Method

### 2.1. Sample

Participants were drawn from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC; Grant et al., 2003a), a NIAAA-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 (Grant et al., 2004, 2005). In brief, informed consent was obtained from all participants. Only those respondents who reported lifetime depressed mood or anhedonia completed the section of the NESARC assessing lifetime occurrence of all DSM-IV symptoms of a MDE, including suicidal ideation and behavior. Of the 43,093 adults surveyed, 1154 endorsed a lifetime history of MDE and manic episode (i.e., bipolar I depression), 494 endorsed a lifetime history of MDE and hypomanic episode (i.e., bipolar II depression), and 5695 endorsed a lifetime history of MDE in the absence of a history of mania or hypomania (i.e., unipolar depression). The present analysis consisted of only those individuals (n=7343; 17% of the total sample).

For participants with bipolar I and bipolar II depression, respectively, average age was 39.4 (SD=14.8) and 35.5 (SD=13.5). Among participants with unipolar depression, average age was 44.7 (SD=15.9). Among participants with bipolar I depression, 63% (n=726) were female, 78% (n=898) were Caucasian, and 83% (n=955) were of non-Hispanic ethnicity. For participants with bipolar II depression, 63% (n=305) were female, 76% (n=375) were Caucasian, and 81% (n=400) were of non-Hispanic ethnicity. Seventy-one percent (n=4018) of participants with unipolar depression were female, 83% (n=4712) were Caucasian, and 84% (n=4774) were of non-Hispanic ethnicity.

## 2.2. Procedure

The Alcohol Use Disorders and Associated Disabilities Interview Schedule-DSM-IV version (AUDADIS-IV; Grant et al., 2001) was used to assess manic, hypomanic, and major depressive episode criteria. Extensive data concerning the psychometric performance of the AUDADIS-IV have been reported elsewhere (Grant et al., 2003b). NESARC estimates of DSM-IV lifetime and 12-month prevalence of illness for BDI were 3.3% and 2.0%, for BDII were 1.1% and 0.8%, and for MDD were 13.2% and 5.3%. These estimates are generally comparable to those found in other recent epidemiological surveys (Kessler et al., 2005b), although it should be noted that the prevalence rates for BDI in the NESARC “slightly exceeded the upper end of the range” of previously reported estimates (Grant et al., 2005, p. 1211). The slightly higher prevalence of BDI in the NESARC may also reflect a cohort effect for BDI that was identified in the National Comorbidity Survey Replication, in which prevalence has been shown to be increasing over time (Kessler et al., 2005a; Parker et al., 2006).

Because the assessment of suicide was limited to the depression module of the AUDADIS-IV and because the current study was focused on suicide within the context of depression across the mood disorders, for purposes of analysis, bipolar I depression was characterized by endorsement of lifetime threshold-level MDE and manic episode criteria, bipolar II depression was characterized by endorsement of lifetime threshold-level MDE and hypomanic episode criteria, and unipolar depression was characterized by endorsement of lifetime threshold-level MDE criteria in the absence of any manic or hypomanic episodes. If respondents endorsed a period of depressed mood and/or anhedonia lasting at least 2 weeks, the AUDADIS-IV was used to assess all remaining DSM-IV depressive symptoms, including 4 items concerning suicidal ideation/attempt: 1) *thoughts of death*, 2) *wanting to*

die, 3) *thoughts of suicide*, and 4) *attempted suicide*. Current study analyses focused on these 4 items.

### 2.3. Data analysis

**Overview**—In order to estimate differential item functioning (DIF) between diagnostic groups, a 2-parameter parametric item response model was used. Item response modeling allows one to examine the likelihood that a particular symptom will be endorsed at a particular level of depression severity (i.e., the latent trait) in different groups. One particular advantage of IRT is that it overcomes the assumption that total number of symptoms determines severity (e.g., 2 symptoms are twice as severe as endorsement of 1 symptom), and instead focuses on the *pattern* of symptom endorsement, relying on a common metric (i.e., standard deviation units) to quantify underlying severity of the latent trait (i.e., depression). As this method ensures that individual characteristics do not affect interpretation of total symptom counts, equal comparisons can be made across groups, anchored on a point along the common metric.

For each item, the 2-parameter model estimates: 1) a severity parameter to describe the point on the latent continuum where a symptom becomes likely to be observed (e.g., > 50%), and 2) a discrimination parameter to describe how rapidly the probability of observing the symptom changes across increasing levels of the latent continuum (e.g., the slope of the item response function).

**2.3.1. Unidimensionality assumption**—The primary assumption of unidimensional item response models is that responses to symptom queries are a function of individual variation along a single underlying dimension. We previously tested this assumption using confirmatory factor analyses (Weinstock et al., 2009, 2010a). In the current study we expand evaluation of each suicide item rather than collapsing responses into a single criterion. To accommodate the expected inter-item associations we evaluated a hierarchical bifactor model (Yung et al., 1999) that models a primary dimension of depression across all depression items and the additional variability within the four suicide items. Confirmatory factor analyses of unidimensional and bifactor models were compared with full-information maximum likelihood estimation ('mirt'; Chalmers, 2012). Comparison of fit statistics indicated that bifactor models provided an improved fit to the data (see Table 1) in respondents with bipolar I and unipolar depression. Although we did not observe a significant difference in model fit within bipolar II respondents, the uni-dimensional model is nevertheless nested within the bifactor model, allowing items to be correlated within their own content groups (Morgan et al., 2015; Reise et al., 2010). Given the nesting of the unidimensional model within the bifactor model, and the advantage of the bifactor model seen in the other larger groups, we proceeded with the same hierarchical model across groups.

**2.3.2. Parametric item response model**—Standard IRT models carry an additional assumption of local independence. That is, symptoms must not be correlated for reasons other than measurement of the latent trait (Lord, 1980). Yet it has been increasingly acknowledged that clinical measures often violate this assumption of local independence by

containing groups of related items, also known as ‘testlets’ (Wainer et al., 2007). Indeed, for the MDE symptoms that assess suicidality, one could reliably predict the presence of one symptom (e.g., thoughts of suicide) from the presence of another (e.g., suicide attempt), irrespective of depressive severity. By expanding the standard IRT model through the addition of a term that represents the variability that is common to this group of items, such local dependence can now be statistically accommodated. This method, based in Testlet Response Theory (TRT; Wainer et al., 2007), allows for the estimation of the relative severity and discrimination of each of the suicide items using a metric representing overall depression severity.

The testlet model we employed (SCORIGHT; Wang et al., 2005) uses Bayesian methods (Gelman et al., 1995) for obtaining estimates of statistical parameters that reflect the severity of and discriminative power of each suicide item. The model also generates a single parameter (theta, where 0 represents the mean value) reflecting each individual's latent continuous level of depressive severity, measured in standard deviation units. Each group is placed on the same metric of depression severity in order to facilitate between-group comparisons of each item after the estimation process (Gelman et al., 1995). This method focuses on obtaining samples from the posterior distribution of each of the model parameters. This is accomplished by using a Markov chain that repeatedly samples for a parameter until it converges on a particular distribution (Markov chain Monte Carlo; MCMC). This posterior distribution then allows for posterior samples to be generated and used to make inferences. For example, in order to estimate the mean value for a (severity or discrimination) parameter, one simply takes the mean value of the posterior sample for that parameter. A primary advantage of MCMC is that these methods do not rely on large-sample theories to estimate standard errors for a parameter (Wang et al., 2005).

**2.3.3. Differential item functioning—**To estimate DIF, we conducted direct comparisons of 10,000 pairs of item parameters drawn at random from the posterior distributions for each item parameter (i.e., severity and discrimination) from each diagnostic group (i.e., bipolar I, bipolar II, and unipolar depression). DIF was determined to be significant if the 95% confidence interval surrounding the average difference score (derived from the distribution of the differences from the 10,000 comparisons) for any given contrast did not include 0 (e.g.,  $p < 0.05$ ). We used a two-parameter model and were most interested in interpreting DIF that occurred in the severity parameter alone, as that would be reflective of the likelihood that a given item will be endorsed at a given severity level. When discrimination parameters are similar, this difference in the likelihood of reporting a symptom is assumed to be uniform across levels of depression severity. However, the discrimination parameter is important in that it can be used to verify that a given item is a good indicator of the latent dimension across respondent groups. When both severity and discrimination parameters are dissimilar, the magnitude of between-group differences may differ across levels of depression severity, and graphic depiction of item response functions can be used to aid interpretation. Yet when discrimination parameters are similar, given that DIF is scaled to standard deviation units, the magnitude of differences in severity parameters can be used to gauge the effect size and can be interpreted using thresholds similar to those used to interpret Cohen's  $d$  (small=0.20, medium=0.50, large=0.80) (Steinberg and Thissen,



2006). DIF was evaluated between bipolar I vs. unipolar depression, bipolar II vs. unipolar depression, and bipolar I vs. bipolar II depression.

### 3. Results

Study results are presented in Table 2. Item response curves depicting study effects are presented in Fig. 1. As reflected in the pattern of raw symptom endorsement (see Table 2), there were overall depression severity differences such that mean depression severity was 0.14 (sd=1.7) standard deviation units higher in BDI versus BDII, 0.24 (sd=1.9) standard deviation units higher in BDI versus MDD, and 0.31 (sd=1.0) standard deviation units higher in BDII versus MDD.<sup>1</sup> These findings underscore the importance of using IRT DIF analyses that account for these overall differences in depression severity. Across all diagnostic groups, endorsement of each suicide item increased as a function of depression severity, and primarily occurred at the higher levels of depression (theta > 0).

Using the *a priori* definition of DIF, direct comparisons of samples from posterior distributions of the severity (b) parameter estimates revealed that those with bipolar I depression were significantly more likely than those with unipolar depression to endorse thoughts of death and suicide attempt. Those with bipolar II depression were significantly more likely than those with unipolar depression to endorse thoughts of death, yet there were no differences between these groups in thoughts of wanting to die, thoughts of suicide, or in attempted suicide. Finally, those with bipolar I and bipolar II depression endorsed thoughts of death, thoughts of suicide, and attempted suicide at similar rates, yet those with bipolar I depression were significantly more likely than those with bipolar II depression to endorse thoughts of wanting to die. However, an inspection of the item response curves in Fig. 1 suggests that effect sizes for the aforementioned group differences were typically small.

Although thoughts of death were significantly more discriminating in bipolar I vs. unipolar depression, the discrimination (a) parameter estimates across diagnostic groups revealed that all of the suicide items, with the exception of thoughts of death, were highly discriminating indices of underlying depression severity.

### 4. Discussion

The aim of this study was to evaluate differences between those with bipolar (I and II) and unipolar depression in the endorsement of suicidal thoughts and attempts. Building upon prior studies, we relied upon testlet response theory models, which allow for the evaluation of response patterns across the underlying latent continuum of depression severity, to evaluate potential diagnostic group differences. The use of a large, population-based sample to evaluate potential differences represents an additional strength of the current study, as it is generalizable to those who may not be treatment-seeking. Consistent with prior research linking depression to suicide (Rihmer and Kiss, 2002; Tondo et al., 2003), endorsement of thoughts of death, thoughts of wanting to die, thoughts of suicide, and attempted suicide all increased as a function of depression severity. Indeed, an inspection of the item response

---

<sup>1</sup>The 'sd' of these means reflects relative variance, such that variance > 1 suggests more variability than was observed in the reference group.

curves (see Fig. 1) revealed that these suicide-related items were rarely endorsed at the lower levels of depression severity (i.e.,  $\theta < 0$ ), and were most frequently endorsed at the highest levels of depression severity (i.e.,  $\theta = 2$ ). This pattern of findings was evident across all three diagnostic groups.

Differences in overall depression severity detected between groups (highest in bipolar I and lowest in unipolar depression) highlight the importance of anchoring comparisons on a point on the underlying latent construct. In doing such, there was some evidence that the bipolar groups were more likely to endorse indices of suicide risk when compared to those with unipolar depression, although patterns of endorsement were somewhat mixed. For example, thoughts of death were endorsed at similar rates in bipolar I and II depression, and were more frequently endorsed in the bipolar groups than in the unipolar group. Yet there was no difference between any of the three groups in the endorsement of thoughts of suicide, arguably a more potent indicator of suicide risk (Gaynes et al., 2004). Contrary to prior arguments that those with bipolar II depression may be at highest risk for suicide (Rihmer and Kiss, 2002), there was no difference between bipolar II and unipolar depression in thoughts of wanting to die, thoughts of suicide, or reported suicide attempt. In fact, those with bipolar I depression were more likely than those with bipolar II depression to endorse thoughts of wanting to die, which has been shown to be another potent indicator of suicide risk (Brown et al., 2005), and only those with bipolar I depression were found to report a greater history of suicide attempt when compared to those with unipolar depression.

Altogether, these data reveal some increased liability for suicidal ideation and attempt in bipolar relative to unipolar depression, and particularly so for bipolar I depression. Yet the effect sizes for such differences, when apparent, were typically quite small (see Table 2 and Fig. 1). The largest effect encountered in this study, falling in the small-to-medium effect size range, was that for the difference in reported suicide attempt between those with bipolar I versus unipolar depression (mean DIF =  $-0.41$ ), indicating that one would expect at least 6% more individuals with bipolar I depression to endorse a suicide attempt than those with unipolar depression. Although such subtle differences may be informative from a public health and empirical perspective, they may not be particularly useful to guide clinical decision making in applied settings. Indeed, irrespective of diagnosis, study data suggest that overall depression severity (vs. specific diagnostic group) may be more important when evaluating risk for suicide among the mood disorders. Further, these data highlight the importance of equating for underlying depression severity in future research evaluating any other potential group differences in risk for suicide. Such conclusions are consistent with more recent studies in clinical samples that failed to find differences between bipolar I, bipolar II, and unipolar depression when overall depression severity was accounted for in the analyses (Holma et al., 2014).

Although not a primary aim of this study, a review of the discrimination (a) parameter estimates (see Table 2) and the accompanying item response curve slopes (see Fig. 1) further revealed that the four suicide items evaluated in this study were, by and large, highly discriminating indicators of underlying depression severity. That is, the probability of endorsement of suicidal ideation and behaviors steeply increased as a function of underlying depression. One exception was for thoughts of death, which was somewhat less



discriminating than the other indicators of suicide risk. These data are consistent with a related analysis that our group conducted, also using NESARC data, which relied upon a larger sample of individuals who had endorsed symptoms of depressed mood or anhedonia, but had not necessarily met criteria for a mood disorder diagnosis (Uebelacker et al., 2010). Although somewhat less discriminating overall, there was some evidence in the current study that thoughts of death were more discriminating in bipolar I than in unipolar depression. These data suggest that the presence of thoughts of death is a somewhat better indicator of underlying depression severity in bipolar I versus unipolar depression; yet, similar to DIF analyses of the severity parameter estimates, this difference was small in magnitude.

When interpreting the above findings, it is important to acknowledge study limitations. First, the data used in this study were cross-sectional, allowing for the evaluation of lifetime suicidal ideation and attempt only. This design feature also limited an ability to account for other clinical course characteristics (e.g., length of illness, rates of mood episode recurrence, hospitalization, medication regimen) that may have potentially influenced patterns of endorsement of suicidal ideation and attempt. Future research, incorporating prospective designs that can more carefully match symptoms and other clinical course features, will be necessary in order to evaluate differences in suicide risk across longitudinal course of the mood disorders. It is also important to note that the NESARC was designed such that questions concerning suicide were evaluated in the depression module of the AUDADIS-IV, thereby limiting an ability to evaluate suicidal ideation and attempt in the context of mania or hypomania, and among individuals with BD who have a lifetime history of mania or hypomania only. Given the evidence that mixed manic and depression symptom presentations may also increase risk for suicide (Tondo et al., 2003), future research that more carefully assesses indices of suicide risk across the full spectrum of mood symptomatology is warranted.

Nevertheless, by addressing several limitations in the existing literature, and by employing a large, well-defined, population-based sample, current study data advance an understanding of suicide risk across the mood disorders in several important ways. Although there was some evidence for greater endorsement of suicidal ideation and attempt in bipolar versus unipolar depression, and perhaps more so in bipolar I depression, findings were not consistent across all indices of ideation and attempt, and differences that did emerge were small in effect. Perhaps more noteworthy was the generally similar pattern of item endorsement across all three diagnostic groups, revealing that suicidal ideation and risk increased dramatically as a function of underlying depression severity. Consistent with the recent emphasis on models of psychopathology that deemphasize DSM-5 categories (Cuthbert and Insel, 2013), study data align with a more dimensional approach to suicide research and treatment.

## Acknowledgments

Contributors

Dr. Weinstock conceptualized the manuscript, managed the literature searches, and wrote the first draft of the manuscript. Dr. Strong provided statistical consultation and conducted the statistical analysis. Drs. Uebelacker and Miller contributed to study conceptualization. All authors contributed to and have approved the final manuscript.

#### Role of the Funding Source

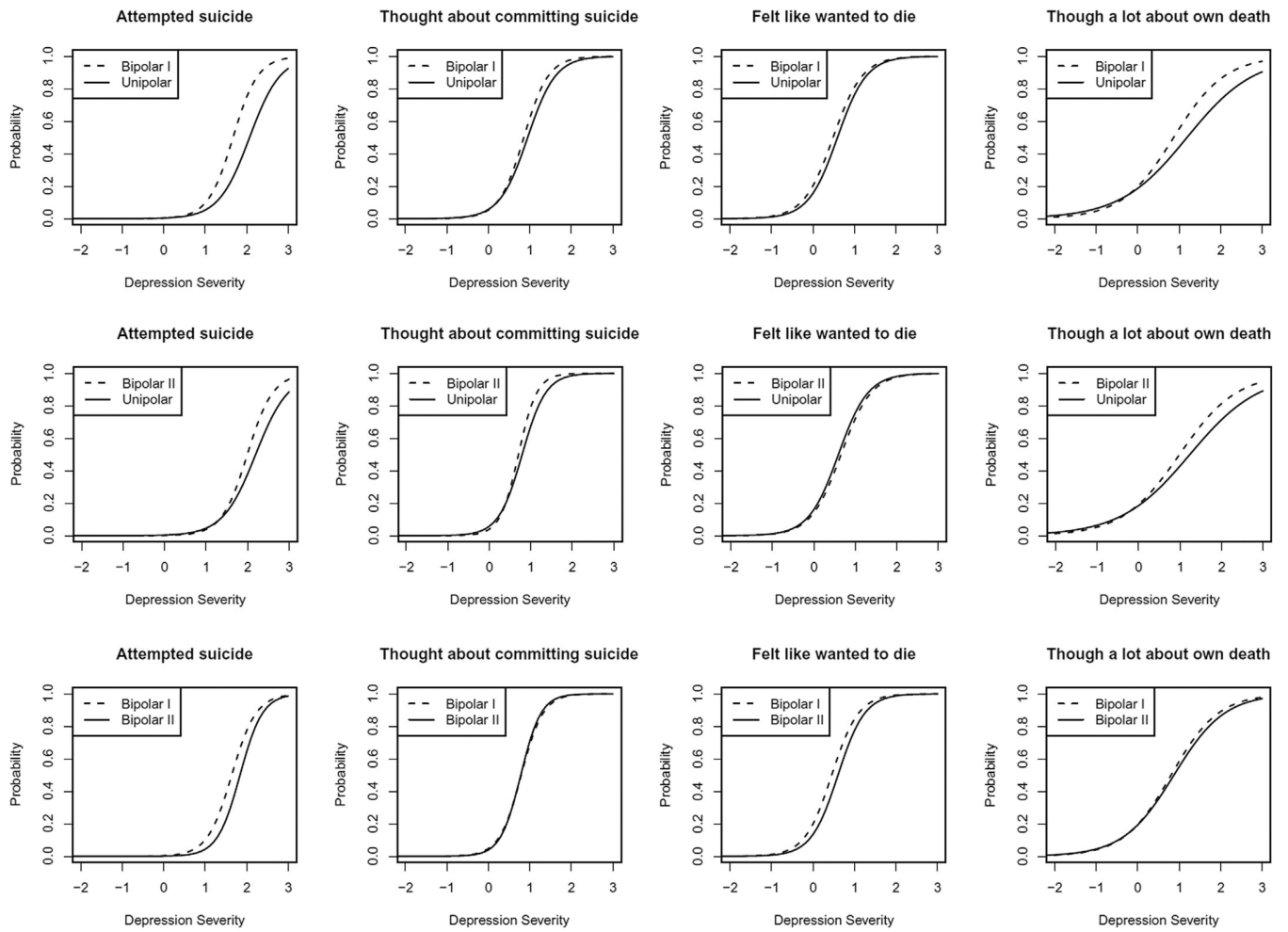
The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) was supported by grants R01AA08159 and K05AA00161 from the National Institute on Alcohol Abuse and Alcoholism. Preparation of this manuscript was supported in part by grant K23MH079907 from the National Institute of Mental Health, awarded to Dr. Weinstock. NIAAA and NIMH had no further role in study design, in the analysis and interpretation of data, in the writing of this report, and the decision to submit the paper for publication.

## References

- Angst F, Stassen HH, Clayton PJ, Angst J. Mortality of patients with mood disorders: follow-up over 34–38 years. *J. Affect Disord.* 2002; 68:167–181. [PubMed: 12063145]
- Baldwin SG, Wainer H, Bradlow ET, Wang X. A review of testlet response theory and its applications. *J. Educ. Behav. Stat.* 2007; 32:333–336.
- Bottlender R, Jager M, Strauss A, Moller HJ. Suicidality in bipolar compared to unipolar depressed inpatients. *Eur. Arch. Psychiatry Clin. Neurosci.* 2000; 250:257–261. [PubMed: 11097169]
- Brown GK, Steer RA, Henriques GR, Beck AT. The internal struggle between the wish to die and the wish to live: a risk factor for suicide. *Am. J. Psychiatry.* 2005; 162:1977–1979. [PubMed: 16199851]
- Cassano GB, Akiskal HS, Savino M, Musetti L, Perugi G. Proposed subtypes of bipolar II and related disorders: with hypomanic episodes (or cyclothymia) and with hyperthymic temperament. *J. Affect Disord.* 1992; 26:127–140. [PubMed: 1447430]
- Chalmers RP. mirt: a multidimensional item response theory package for the R environment. *J. Stat. Softw.* 2012; 48:1–29.
- Chen YW, Dilsaver SC. Lifetime rates of suicide attempts among subjects with bipolar and unipolar disorders relative to subjects with other axis I disorders. *Biol. Psychiatry.* 1996; 39:896–899. [PubMed: 8860192]
- Coryell W, Andreasen NC, Endicott J, Keller M. The significance of past mania or hypomania in the course and outcome of major depression. *Am. J. Psychiatry.* 1987; 144:309–315. [PubMed: 3826429]
- Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* 2013; 11:126. [PubMed: 23672542]
- Endicott J, Nee J, Andreasen N, Clayton P, Keller M, Coryell W. Bipolar II. Combine or keep separate? *J. Affect Disord.* 1985; 8:17–28. [PubMed: 3156908]
- Fiedorowicz JG, Leon AC, Keller MB, Solomon DA, Rice JP, Coryell WH. Do risk factors for suicidal behavior differ by affective disorder polarity? *Psychol. Med.* 2009; 39:763–771. [PubMed: 18667100]
- Gaynes BN, West SL, Ford CA, Frame P, Klein J, Lohr KN. Screening for suicide risk in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* 2004; 140:822–835. [PubMed: 15148072]
- Gelman, A.; Carlin, J.; Stern, H.; Rubin, D. *Bayesian Data Analysis.* Chapman and Hall; London: 1995.
- Grant, BF.; Dawson, DA.; Hasin, DS. *The Alcohol Use Disorder and Associated Difficulties Interview Schedule-DSM-IV Version.* National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2001.
- 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.* 2003a; 71:7–16. [PubMed: 12821201]
- Grant, BF.; Kaplan, KD.; Shepard, J.; Moore, T. *Source and Accuracy Statement for Wave 1 of the 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions.* National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2003b.

- 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]
- Grant BF, Stinson FS, Hasin DS, Dawson DA, Chou SP, Ruan WJ, Huang B. 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]
- Holma KM, Haukka J, Suominen K, Valtonen HM, Mantere O, Melartin TK, Sokero TP, Oquendo MA, Isometsa ET. Differences in incidence of suicide attempts between bipolar I and II disorders and major depressive disorder. *Bipolar Disord.* 2014; 16:652–661. [PubMed: 24636453]
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry.* 2005a; 62:593–602. [PubMed: 15939837]
- 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.* 2005b; 62:617–627. [PubMed: 15939839]
- Lester D. Suicidal behavior in bipolar and unipolar affective disorders: a meta-analysis. *J. Affect Disord.* 1993; 27:117–121. [PubMed: 8440807]
- Lord, F. Applications of Item Response Theory to Practical Testing Problems. Erlbaum; Hillsdale, NJ: 1980.
- Moreno C, Hasin DS, Arango C, Oquendo MA, Vieta E, Liu S, Grant BF, Blanco C. Depression in bipolar disorder versus major depressive disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Bipolar Disord.* 2012; 14:271–282. [PubMed: 22548900]
- Moreno DH, Andrade LH. The lifetime prevalence, health services utilization and risk of suicide of bipolar spectrum subjects, including subthreshold categories in the Sao Paulo ECA study. *J. Affect Disord.* 2005; 87:231–241. [PubMed: 16009430]
- Morgan G, Hodge K, Wells K, Watkins M. Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via monte carlo simulations. *J. Intell.* 2015; 3:2–20.
- Parker G, Brotchie H, Fletcher K. An increased proportional representation of bipolar disorder in younger depressed patients: analysis of two clinical databases. *J. Affect Disord.* 2006; 95:141–144. [PubMed: 16777233]
- Raja M, Azzoni A. Suicide attempts: differences between unipolar and bipolar patients and among groups with different lethality risk. *J. Affect Disord.* 2004; 82:437–442. [PubMed: 15555695]
- Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *J. Pers. Assess.* 2010; 92:544–559. [PubMed: 20954056]
- Rihmer Z, Kiss K. Bipolar disorders and suicidal behaviour. *Bipolar Disord.* 2002; 4(Suppl. 1):S21–S25.
- Steinberg L, Thissen D. Using effect sizes for research reporting: examples using item response theory to analyze differential item functioning. *Psychol. Methods.* 2006; 11:402–415. [PubMed: 17154754]
- 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*. Lawrence Erlbaum Associates; Hillsdale, NJ: 1993. p. 67-113.
- Tondo L, Baldessarini RJ, Hennen J, Minnai GP, Salis P, Scamonatti L, Masia M, Ghiani C, Mannu P. Suicide attempts in major affective disorder patients with comorbid substance use disorders. *J. Clin. Psychiatry.* 1999; 60(Suppl. 2):S63–S66. 113–116.
- Tondo L, Isacson G, Baldessarini R. Suicidal behaviour in bipolar disorder: risk and prevention. *CNS Drugs.* 2003; 17:491–511. [PubMed: 12751919]
- Tondo L, Lepri B, Baldessarini RJ. Suicidal risks among 2826 Sardinian major affective disorder patients. *Acta Psychiatr. Scand.* 2007; 116:419–428. [PubMed: 17997721]

- Uebelacker LA, Strong D, Weinstock LM, Miller IW. Likelihood of suicidality at varying levels of depression severity: a re-analysis of NESARC data. *Suicide Life Threat Behav.* 2010; 40:620–627. [PubMed: 21198331]
- Vieta E, Benabarre A, Colom F, Gasto C, Nieto E, Otero A, Vallejo J. Suicidal behavior in bipolar I and bipolar II disorder. *J. Nerv. Ment. Dis.* 1997; 185:407–409. [PubMed: 9205428]
- Wainer, H.; Bradlow, ET.; Wang, X. *Testlet Response Theory and its Applications.* Cambridge University Press; New York: 2007.
- Wang, X.; Bradlow, ET.; Wainer, H. *User's guide for SCORIGHT (Version 3.0). A computer program for scoring tests built of testlets including a module for covariate analysis (Research Report 04–09).* Educational Testing Services; Princeton, NJ: 2005.
- Weinstock LM, Strong D, Uebelacker LA, Miller IW. Differential item functioning of DSM-IV depressive symptoms in individuals with a history of mania versus those without: An item response theory analysis. *Bipolar Disord.* 2009; 11:289–297. [PubMed: 19419386]
- Weinstock LM, Strong D, Uebelacker LA, Miller IW. DSM-IV depressive symptom expression among individuals with a history of hypomania: a comparison to those with or without a history of mania. *J. Psychiatry Res.* 2010a; 44:979–985.
- Weinstock LM, Strong D, Uebelacker LA, Miller IW. Use of item response theory methodology to evaluate differences between bipolar and unipolar depression. *US Psych.* 2010b; 3:15–20.
- Weinstock LM, Strong D, Uebelacker LA, Miller IW. Differences in depression symptom endorsement between bipolar disorder and major depressive disorder: lessons learned from the National Epidemiologic Survey on Alcohol and Related Conditions. *Bipolar Disord.* 2013; 15:110–111. [PubMed: 23231053]
- Yung YF, Thissen D, McLeod LD. On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika.* 1999; 64:113–128.
- Zalsman G, Braun M, Arendt M, Grunebaum MF, Sher L, Burke AK, Brent DA, Chaudhury SR, Mann JJ, Oquendo MA. A comparison of the medical lethality of suicide attempts in bipolar and major depressive disorders. *Bipolar Disord.* 2006; 8:558–565. [PubMed: 17042829]



**Fig. 1.** Differential item functioning (DIF) of suicide items by diagnostic group.

**Table 1**

Comparison of fit from confirmatory unidimensional and hierarchical bifactor models within bipolar I, bipolar II, and unipolar respondents.

	<b>AIC</b>	<b>BIC</b>	<b>logLik</b>	$\chi^2$	<b>df</b>	<b><i>p</i></b>
Bipolar I ( <i>n</i> = 1154)						
Single dimension	10,034.968	10,135.988	-4997.484			
Hierarchical bifactor	9710.836	9862.366	-4825.418	344.132	10	< 0.001
Bipolar II ( <i>n</i> = 494)						
Single dimension	4235.604	4319.655	-2097.802			
Hierarchical bifactor	4247.236	4373.312	-2093.618	8.368	10	0.593
Unipolar ( <i>n</i> = 5695)						
Single dimension	51,595.56	51,728.51	-25,777.78			
Hierarchical bifactor	51,524.63	51,724.05	-25,732.32	90.934	10	< 0.001

*Note.* AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; logLik = Log-Likelihood.



Unadjusted frequencies of endorsement, discrimination (a) and severity (b) parameter estimates, and differential item functioning (DIF) of suicide items between bipolar I, bipolar II, and unipolar depression.

**Table 2**

Item	%	Contrast	a	SE (a)	Mean DIF	95% CI	b	SE (b)	Mean DIF	95% CI
Thoughts of death	50	Bipolar I	1.62	0.17	0.35*	[0.10, 0.62]	0.87	0.08	-0.34*	[-0.48, -0.21]
	32	Unipolar	1.24	0.10			1.19	0.09		
Wanting to die	63	Bipolar I	2.83	0.34	-0.03	[-0.66, 0.75]	0.48	0.05	-0.11	[-0.22, 0.00]
	46	Unipolar	2.86	0.21			0.59	0.03		
Thoughts of suicide	54	Bipolar I	3.38	0.39	0.41	[-0.32, 1.28]	0.85	0.05	-0.09	[-0.20, 0.01]
	36	Unipolar	2.97	0.19			0.95	0.04		
Attempted suicide	25	Bipolar I	3.34	0.45	0.67	[-0.17, 1.73]	1.67	0.07	-0.41*	[-0.59, -0.22]
	9	Unipolar	2.70	0.18			2.07	0.08		
<hr/>										
Thoughts of death	41	Bipolar II	1.45	0.22	0.23	[-0.06, 0.56]	0.99	0.13	-0.27*	[-0.51, -0.03]
	32	Unipolar	1.20	0.13			1.24	0.13		
Wanting to die	51	Bipolar II	2.75	0.45	-0.06	[-0.92, 1.07]	0.65	0.07	0.06	[-0.09, 0.22]
	46	Unipolar	2.80	0.21			0.59	0.03		
Thoughts of suicide	44	Bipolar II	4.53	0.86	0.98	[-0.35, 2.65]	0.70	0.06	-0.09	[-0.18, 0.01]
	36	Unipolar	3.53	0.26			0.79	0.04		
Attempted suicide	15	Bipolar II	3.25	0.72	0.74	[-0.28, 2.37]	2.00	0.16	-0.22	[-0.50, 0.08]
	9	Unipolar	2.54	0.24			2.19	0.16		
<hr/>										
Thoughts of death	50	Bipolar I	1.76	0.10	0.13	[-0.34, 0.59]	0.80	0.05	-0.07	[-0.25, 0.11]
	41	Bipolar II	1.64	0.19			0.87	0.07		
Wanting to die	63	Bipolar I	3.09	0.47	-0.02	[-1.21, 1.23]	0.45	0.07	-0.14*	[-0.26, -0.01]
	51	Bipolar II	3.10	0.57			0.60	0.09		
Thoughts of suicide	54	Bipolar I	3.80	0.53	-0.43	[-3.15, 1.27]	0.79	0.06	0.01	[-0.13, 0.15]
	44	Bipolar II	4.10	0.86			0.78	0.07		
Attempted suicide	25	Bipolar I	3.42	0.45	-0.21	[-2.15, 1.41]	1.64	0.06	-0.20	[-0.46, 0.01]
	15	Bipolar II	3.69	0.75			1.84	0.10		

Note. Mean DIF = average difference between item parameters drawn from posterior distributions, with 95% confidence intervals. DIF was determined to be significant if the 95% confidence interval surrounding the mean DIF score did not include 0.

*p* < .05  
\*

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript