International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 17(4): 198–209 (2008)
Published online 12 September 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.262

Development of the Bipolar Inventory of Symptoms Scale: concurrent validity, discriminant validity and retest reliability

JODI M. GONZALEZ, CHARLES L. BOWDEN, MARTIN M. KATZ, PETER THOMPSON, VIVEK SINGH, THOMAS J. PRIHODA, MARTHA DAHL

Department of Psychiatry, UT Health Science Center, San Antonio, TX, USA

Abstract

Scales used in studies of bipolar disorder have generally been standardized with major depressive or hospitalized manic patients. A clinician rated scale based on a semi-structured interview for persons with bipolar disorder, with comprehensive coverage of bipolar symptomatology, is needed. We report concurrent, divergent and convergent psychometric reliability, discriminant validity and relationship to a measure of overall function for a new psychometric rating instrument. A primarily outpatient sample of 224 subjects was assessed using the Bipolar Inventory of Symptoms Scale (BISS). The BISS total score and depression and mania subscales were compared to the Young Mania Rating Scale (YMRS), the Montgomery Asberg Depression Rating Scale (MADRS) and the Global Assessment of Functioning Scale (GAF). Clinical mood states were also compared using the BISS. The BISS scores demonstrated good concurrent validity, with estimates (Pearson correlations) ranging from 0.74 to 0.94 for YMRS and MADRS and test–retest reliability from 0.95 to 0.98. BISS concurrent validity with the GAF was significant for four clinical states, but not mixed states. The BISS discriminated primary bipolar mood states as well as subjects recovered for eight weeks compared to healthy controls. In conclusion, the BISS is a reliable and valid instrument broadly applicable in clinical research to assess the comprehensive domains of bipolar disorder. Future directions include factor analysis and sensitivity to change from treatment studies. Copyright © 2008 John Wiley & Sons, Ltd.

Key words: psychiatric status rating scales, psychometrics, bipolar disorder, psychological assessment

Introduction

Because of the complex phenomenology of bipolar disorders (BDs) (Cassidy et al., 1998; Dennehy et al., 2004), rating scales are particularly important tools in treatment trials and pathophysiological studies. Traditional clinician administered depression scales measure classic forms of depression and have generally good reliability and validity (Bagby et al., 2004). Only one instrument has been standardized and validated separately for bipolar depression, two forms of the Inventory of Depressive Symptoms [Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) 30 items and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) 16 items] clinician rated scales (Bagby et al., 2004; Trivedi et al., 2004). Several mania rating scales have been developed, but are limited by the small number of items and the narrow range of severity of strictly hospitalized patients on which they were psychometrically developed (Young et al., 1978). Consequent to evidence that the behavioral disturbances in BD extend beyond narrowly defined depressive and manic symptoms, there is a growing need for a bipolar rating scale which measures the spectrum of symptoms seen in BD (Berk et al., 2004).

The IDS-C and QIDS-C were utilized as outcome measures in a large study (Texas Medication Algorithm Project) of BD, major depression and schizophrenic/schizoaffective persons. Analyses with a subset of the BD population showed that the IDS-C and QIDS-C had good sensitivity to change with treatments. Change in the IDS was associated with changes in a measure of general functioning, the Short Form-12 (Trivedi et al., 2004). The same group developed a 10 item scale principally derived from the Brief Psychiatric Rating Scale (BPRS), called the Brief Bipolar Disorder Symptom Scale (BDSS), to provide a short scale on overall psychopathology, which performed similarly to the BPRS (Dennehy et al., 2004). In terms of relationship to a mania rating instrument [Clinician Administered Rating Scale (CARS)], the correlation was 0.25 and to the depressive symptoms scale (IDS-C) the correlation was 0.21. The full BPRS score correlations with the CARS (0.23) and the depression scale (0.22) were similar.

We developed the Bipolar Inventory of Symptoms Scale (BISS) to address the need for a structured interview scale that comprehensively and adequately captures the full spectrum of bipolar symptomatology and related objectives, e.g. sensitivity to changes in symptoms in the course of illness and treatment and analyses of fundamental behavioral domains/components of the disorder. Research over the past few decades has uncovered characteristics of the syndrome not usually included in rating scales, such as sharpened thinking, elevated evening energy, impulsivity, risky behavior and affective lability. The BISS includes these items which in a preliminary study demonstrated ability to demonstrate discriminate mood states (Bowden et al., 2007).

In this report we provide concurrent validity of the BISS for constructs of depression and mania through comparison of established rating scales and corresponding measures from the BISS. Additionally, we examined the relationship of BISS scores to an overall measure of functioning, the Global Assessment of Functioning Scale (GAF). To test whether BISS scores were specific indices of depression and mania, rather than non-specific burden of psychopathology, we also examined divergent validity. We conducted analyses to determine if BISS scores discriminated between bipolar clinical mood states. We predicted that validity would be supported by significant differentiation between several clinical states. That is, the BISS would be able to distinguish between clinical mood states of depression by significantly higher scores on depression subscales as compared to manic/hypomanic mood states, recovered status and healthy controls. We predicted the mania subscale would be able to distinguish from persons in a depressed mood state, those recovered, and healthy controls. For mixed mood state subjects, we predicted BISS depression scale scores would not differ significantly from a score of depressed subjects, and that manic scores would not differ significantly from scores of manic/hypomanic subjects. Finally we examined retest reliability of the BISS across raters for a subset of subjects.

Methods

Procedures

The BISS items were adapted from published rating scales, particularly the Schedule of Affective Disorders and Schizophrenia – clinical rating (SADS-C, Endicott and Spitzer, 1978), and new items developed based on all relevant available literature to cover the full spectrum of bipolar symptomatology and affective disturbances. The BISS is comprised of 44 items; 22 items for the depression scale and 22 items for the mania scale. The BISS is administered utilizing semi-structured questions, with all items rated on a five point Likert-type scale, 0–4, with each severity score operationally defined. Examples of the structure of items are shown in Table 1. Ratings are based on the most recent seven day period, utilizing information from self report, and, when available, family and clinician observation both outside and during the interview. For additional information on scale objectives, development and methods, see Bowden et al. (2007).

Subjects were recruited by clinic referral, including those participating in ongoing clinical trials, and flier advertisements. Healthy control subjects were recruited by flier advertisement within the medical center as well as a local trade school. After procedures were fully explained, written informed consent was obtained on all subjects. The interview included a semi-structured diagnostic interview, demographic questions and the assessment set of instruments. Participants were first assessed with the BISS, followed by the Young Mania Rating Scale (YMRS) and the Montgomery Asberg Depression Rating Scale (MADRS), which were completed by the same rater (see later for further description of measures).

DOI: 10.1002/mpr

Table 1. Item description (#2 from depression subscale, #28 from mania subscale)

2. Sadness-observed: Appears despondent, gloomy, despairing, as reflected in speech, facial expression and posture. Rate also by inability to brighten up. **Have others commented that you appeared sad, blue, or unhappy?**	0 Not at all 1 Slight; e.g. looks dispirited, but brightens up easily 2 Mild; e.g. more physical manifestations of sadness, and less response to positive stimulus 3 Moderate; e.g. often appears sad and unhappy 4 Severe; e.g. extreme and continuous gloom and despondency
28. Increased social interest: Increased interest in, attention to others. Associated with increased social interactions and time spent with others. **Have you had more interest than usual in spending time with other people, including church, family, friends, co-workers?** *Did you actually spend more time interacting with others? How much more time than usual for you does this represent? Did you make social contact with people whom you do not know, or have not seen in a long time? What about talking on the phone, or using the internet for conversations?*	0 Not at all 1 Slight increase in interest in social interaction. 2 Mild; actual increase in social interaction, associated with actively seeking out others. 3 Moderate; substantially increased time spent with others, causing minor impairment, e.g. inappropriate renewing of old acquaintances. 4 Marked increase in social interactions, associated with intrusive, unwanted interactions, contacts strangers.

Note: bold questions required, italics required if positive responses to bold questions.

Test–retest sample

Twenty-one of the 224 subjects were retested within three days of the initial BISS administration. Recruitment for this arm of the study had an aim of 20 subjects to provide power of at least 0.80 to detect correlations of 0.57 with two-tailed tests at $p \leq 0.05$ (Cohen, 1988). Recruitment for the test–retest subjects was complete when 20 subjects were tested. The retest assessment was completed by a second rater who was blind to the scores from the first BISS administration. Subjects had varied diagnoses; 13 with bipolar I disorder, four bipolar II disorder and four with major depression.

Missing items

For the BISS scores, five subjects had one missing item, one had two and one subject had three missing items, resulting in an acceptable missing item rate of 0.1%. When items were missing, the average rating for the subscale (depression or mania) was used to replace the missing score. On MADRS scores, no subjects for which MADRS data were collected (214/224 had MADRS data) had missing item data. For the YMRS, no subjects for which YMRS data were collected had missing item data (n = 196/224 had YMRS data).

Raters

Ten raters participated in this validity and reliability phase. Four were psychiatrists, two psychologists, and one each was a psychology resident, a foreign graduate medical student, a registered nurse and a psychiatric social worker. Seven of the raters were involved in the development of the measure. The three clinician raters trained observed at least one BISS administration while also rating concurrently. Scores were compared and any discrepancy in ratings was discussed. A trained rater then observed one BISS administered by each of the three raters. No other criteria were used to determine rater certification.

Measures

To establish diagnosis, raters completed the Mini International Neuropsychiatric Interview (MINI Plus Version 5.0), a semi-structured interview designed to yield lifetime and current Axis I Disorder diagnoses (Sheehan et al., 1998) based on the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV). The MINI Plus has well established reliability and validity (Sheehan et al., 1998), and has been widely applied as the diagnostic instrument or record

DOI: 10.1002/mpr

in major recent studies of BD, including the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD; Sachs et al., 2003; Perlis et al., 2004). BD subjects met lifetime criteria for any hypomanic or manic episode. Major depressive disorder subjects met lifetime criteria for a major depressive episode with no history of mania or hypomania. Inclusion criteria for controls were no lifetime history of psychiatric comorbidity in major Axis I Disorders (mood disorder, anxiety disorder, history of psychosis, substance dependence disorders, eating disorders), current (within the last 12 months) or lifetime (Sheehan et al., 1998). For Anxiety Disorder, current as well as lifetime disorders had to be negative (panic disorder, agoraphobia, social phobia, obsessive compulsive disorder, and post-traumatic stress disorder). Alcohol Comorbidity included abuse and/or dependence, Substance Use Comorbidity included current abuse and/or dependence of any non-alcohol substance. Participants meeting criteria only for lifetime alcohol or substance abuse were not excluded from the healthy control group. For Eating Disorders, participants had no history of anorexia nervosa or bulimia nervosa.

For this report, clinical status was categorized as depression, mixed state, hypomania or mania, subsyndromal, or recovered. To determine mood episode status, two methods were used to establish the nature of the current episode (one method was used per subject). First, a semi-structured interview form of DSM-IV clinical syndromes (Sachs, 1990) utilized in the STEP-BD, was conducted at the time of the BISS interview ($N = 128$). The subsyndromal category was defined as those persons not meeting recovered criteria (≤2 clinically significant mood symptoms for at least eight weeks) but not meeting DSM-IV clinical episode criteria. Second, for subjects enrolled outside of the STEP-BD, YMRS (version 2.1) and the MADRS (version 1.0) scores were used to determine mood state ($N = 96$), both which assess symptoms from the prior week. The MADRS has been used with similar cutoff scores to determine mood episode versus no mood episode (Kasper et al., 2006; Loo et al., 2007; Thase et al., 2006) and is used often as a primary outcome measure in clinical trials. To meet criteria for mania, a YMRS total score >17 was required; for hypomania a score of 12–17. Hypomania derived scores using the YMRS are often utilized in BD treatment studies (Altshuler et al., 2006; Frye et al., 2007; Schaffer et al., 2006). Scores from 9–11 were classed as subsyndromal and scores below 9 were classed as recovered. For depression, MADRS total scores >17 were defined as a depressive episode, 9–17 for subsyndromal, and <9 was defined as recovered. The lower and upper bounds for subsyndromal were adapted from recent studies which have utilized similar bounds to define subsyndromal illness states in BDs (Frye et al., 2006). For mixed episode classification, both the depressive episode threshold (>17) and at least the hypomania episode threshold (≥12) had to be met. We included hypomanic mixed states because of their consistent high proportion in samples of BD (Akiskal and Benazzi, 2005; Benazzi et al., 2004; Bennazi, 2004).

The YMRS and MADRS items were used to examine the relationship of the BISS to depressive and manic symptoms. The GAF was used as an overall measure of psychological, social and occupational functioning, a commonly used measure of functioning on a hypothetical continuum of mental health-illness (Parabiaghi et al., 2006). Inter-rater reliability among clinicians and researchers ranges from 0.54 to 0.87 with researcher ratings in a recent study ranging from 0.81 to 0.85 (Vatnaland et al., 2007). The GAF scores were rated on the subject's status for the previous month, assessed at the time of the BISS and rating scale assessments.

Statistical analysis

For demographic data and baseline group comparisons reported, chi square and analysis of variance were utilized. For determining the relationship of the BISS to established methods, BISS scores were correlated with relevant rating scales (MADRS, YMRS and GAF). Pearson product moment correlations were utilized. For test–retest analyses, intra-class correlation coefficients were calculated. Analysis of variance was used for across group comparisons of clinical status and BISS total and subscale scores. For each ANOVA, residual plots were examined to verify the assumptions of the ANOVA. Standard square root transformations were used when indicated to better satisfy ANOVA assumptions of equal distribution and homogeneity of variance. Bonferroni corrections were used to adjust for multiple comparisons when any omnibus F was statistically significant at $p < 0.05$; significance testing was two-tailed with p values set at ≤0.05 to indicate significant statistical comparisons.

DOI: 10.1002/mpr

Results

Sample characteristics

A total of 232 patients were consented to participate in the study. Eight subjects who consented were not included in analyses due either to not qualifying as a healthy control or having a diagnosis of BP NOS or Schizoaffective Disorder, leaving a total sample of 224. Participants included 141 subjects with bipolar I disorder, 42 with bipolar II disorder, 24 with major depression, and 17 healthy control subjects (Table 2). Participants ranged in age from 18 to 80 years old, and were 37.6% ethnic/racial minority backgrounds (N = 89). Of the Hispanic participants, 88% were of Mexican/Mexican American origin, none reported Puerto Rican or Cuban descent and 12% were 'other Hispanic'

Table 2. Demographics of validity sample (N = 224)

	Patient subjects (N = 207) Frequency[1]	Healthy controls (N = 17) Frequency[1]
Age – mean (standard deviation)		
Patient population – 40.06 (12.92) 18–80 years old		
Healthy controls – 31.47 (12.02) 18–57 years old		
Gender		
Female	121 (58.5)	11 (64.7)
Male	86 (41.5)	6 (35.3)
Education[2]		
<High school	16 (9.2)	2 (11.8)
High school	72 (41.4)	5 (29.4)
Some college	19 (10.9)	3 (17.6)
Bachelors degree and up	67 (38.5)	7 (41.2)
Marital status[2]		
Single never married	52 (29.2)	8 (47.1)
Married/cohabiting	73 (41.0)	8 (47.1)
Disrupted marriage	53 (29.8)	1 (5.9)
Ethnicity/race[2]		
African American	21 (10.3)	2 (12.5)
Caucasian	123 (60.6)	7 (43.8)
Hispanic	52 (25.6)	6 (37.5)
Other	7 (3.5)	1 (6.3)
Clinical status		
Depressed	64 (30.9)	n/a
Mixed	48 (23.2)	
Manic/hypomanic	34 (16.4)	
Subsyndromal	14 (6.8)	
Recovered	47 (22.7)	
Diagnosis		
Bipolar I disorder	141 (68.1)	
Bipolar II disorder	42 (20.3)	
Major depressive disorder	24 (11.6)	
Healthy control		17 (100)

[1] Percentages are shown in parentheses.
[2] Some data are missing.
Note: Independent observations; n/a, not available.

DOI: 10.1002/mpr

background. Ninety-one percent of Hispanic/Latino subjects were US born. Participants were asked in what language they preferred to speak. Two of the participants had no preference for speaking English or Spanish; three preferred to speak in Spanish but reported fluency in English. Participants ranged in education level from less than eighth grade to a baccalaureate degree and above, with approximately 50% of the sample reporting a high school education or less. Table 2 provides further description of the sample. Clinical status groups for which analyses of variance were performed (depressed, manic/hypomanic, mixed, recovered, subsyndromal and healthy controls) did not differ on gender, marital status, education level, ethnicity/race or rater conducting the interview (data not shown).

Novel items for bipolar rating scales included in the BISS each correlated significantly to the BISS total scale using Pearson correlations: sharpened thinking $r = 0.19$, elevated evening energy $r = 0.37$, impulsivity $r = 0.58$, risky behavior $r = 0.44$ and affective lability $r = 0.66$ (at a p value <0.01).

Concurrent validity

We compared BISS total scores for all subjects, independent of mood state, with total scores on the MADRS, YMRS and the GAF (Table 3). Total BISS scores correlated significantly with the MADRS (0.81, $p < 0.001$), the YMRS (0.74, $p < 0.001$) and the GAF (−0.76, $p < 0.001$). BISS depression subscale correlations with the MADRS were 0.94 for all subjects combined and 0.81 for recovered subjects. For subjects in depressive episodes, the BISS depression subscale score correlation with the MADRS was 0.83 ($p < 0.001$). BISS mania subscale correlations with the YMRS were 0.91 across all mood states, and 0.84 for recovered patients. For manic/hypomanic subjects, the BISS mania subscale correlation with the YMRS was 0.78 ($p < 0.001$) (Table 3).

We investigated whether BISS scores within mood state groups would correlate with overall functioning, as measured by the GAF score. For recovered subjects, correlations of BISS total score and the two subscales with GAF ranged from −0.48 to −0.69. The total sample

Table 3. Concurrent and divergent validity for BISS total score, depression subscale and mania subscale

Rating scale	Total sample ($N = 224$)[1]	Current depressive episode[1] ($N = 64$)	Currently manic/hypomanic ($N = 34$)	Current mixed episode ($N = 48$)	Recovered ($N = 47$)
BISS total scale					
MADRS	*0.81****	0.64***	0.68***	0.67***	0.65***
YMRS	*0.74****	0.69***	0.41**	0.66***	0.73***
GAF	*−0.76****	*−0.21*	*−0.35**	*0.06*	*−0.69****
Depression subscale					
MADRS	*0.94****	*0.83****	0.90***	*0.77****	0.81***
YMRS	**0.38*****	**0.26**	**−0.14**	**0.30**	**0.26**
GAF	−0.68***	−0.35**	0.12	0.03	−0.54***
Mania subscale					
MADRS	**0.31*****	**0.11**	**−0.06**	**0.31***	**0.16**
YMRS	0.91***	0.81***	*0.78****	*0.79****	0.84***
GAF	−0.56***	0.06	−0.61***	0.06	−0.48***

[1] Total sample correlations include control subjects and subsyndromal subjects (other categories do not include):

- MADRS, Mongtomery Asberg Depression Scale; YMRS, Young Mania Rating Scale; GAF, Global Assessment of Functioning Scale.
- Depressive episode included unipolar and bipolar depression.
- Pearson product moment correlations.

Note: Italic typeface represents concurrent validity outcomes of interest; bold typeface represents divergent validity.
* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

DOI: 10.1002/mpr

also revealed strong correlations ranging from −0.56 to −0.76. Depressed subjects had moderate but significant correlations between BISS depression subscale scores and GAF (−0.35). Similarly, for manic/hypomanic subjects BISS mania subscale scores were significantly correlated with GAF scores (−0.61). For mixed episode subjects, none of the three BISS scores were significantly correlated with GAF scores.

We examined divergent validity in the total sample and in the clinical episode subsamples by comparing the BISS depression scale correlations with the YMRS and the BISS mania subscale correlations with the MADRS. The BISS depression subscale correlations with the YMRS were low and generally non-significant across samples (0.38 to −0.14) as were the BISS mania subscale correlations with the MADRS (0.31 to −0.06). The contrasts in coefficients are shown clearly within the depressive, manic, and mixed episode subsample analyses (Table 3).

Discriminant validity

We conducted three analyses of variance to determine if clinical groups and the healthy controls could be distinguished by BISS total and subscale scores (see Table 4). For the three ANOVAs, the square root transformation was applied as the residual plots indicated that the transformation best satisfied assumptions of the ANOVA. Raw data means are reported for the descriptive statistics; multiple comparisons and *F*-tests are reported for the transformed data. The overall *F* for the BISS total score was significant [F (5,218) = 139.12, $p < 0.001$]. The BISS total score distinguished between all episode types (depressed, mixed or manic/hypomanic) and subsyndromal, recovered, and healthy controls, using Bonferroni *post hoc* comparisons (Table 4). Subjects in a depressive episode did not differ on BISS total score from those in a manic/hypomanic episode.

The overall *F* for the BISS depression subscale was significant [F (5,218) = 125.57, $p < 0.001$], with similar

Table 4. Discriminant validity results for the BISS total score, depression and mania subscales

Clinical mood state	Mean	Standard deviation	Pairwise comparisons significant compared to clinical mood state
*BISS total score****			
Depressive episode	54.11	15.94	Mixed, subsyndromal, recovered, control (not manic)
Mixed episode	72.35	17.23	All
Manic/hypomanic	57.82	15.52	Mixed, subsyndromal, recovered, control (not depressed)
Subsyndromal	38.43	16.31	All
Recovered	16.92	9.37	All
Control	5.89	4.30	All
*BISS depression subscale****			
Depressive episode	41.64	10.92	Manic, subsyndromal, recovered, control (not mixed)
Mixed episode	44.10	10.82	Manic, subsyndromal, recovered, control (not depressed)
Manic/hypomanic	24.03	12.51	Depressed, mixed, recovered, control (not subsyndromal)
Subsyndromal	24.50	13.12	Depressed, mixed, recovered, control (not manic)
Recovered	10.02	6.32	All
Control	3.76	3.56	All
*BISS mania subscale****			
Depressive episode	12.47	10.01	Mixed, manic, recovered control (not subsyndromal)
Mixed episode	28.25	10.52	Depressed, continued, recovered, control (not manic)
Manic/hypomanic	33.79	11.01	Depressed, subsyndromal, recovered, control (not mixed)
Subsyndromal	13.93	6.40	mixed, manic, recovered, control (not depressed)
Recovered	6.89	5.85	All
Control	2.12	1.69	All

*** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$, analyses of variance utilized with *post hoc* Bonferroni corrections.

DOI: 10.1002/mpr

findings to the BISS total scores, except that pairwise comparisons did not distinguish depressed and mixed episode subjects, as both scored high on the depression subscale (see Table 4). The mania subscale overall F was significant [F (5,218) = 64.89, $p < 0.001$], also with a similar pattern of findings and good discrimination between mood states. Pairwise comparisons did not distinguish between mania/hypomania group and the mixed episode group, as both scored high on the mania subscale.

Recovered subjects versus healthy controls

In pairwise comparisons, the three analyses of variance for BISS total score, BISS mania scale and BISS depression scale distinguished between recovered subjects with BD and healthy controls. The mean (standard deviation) BISS total score for recovered subjects was 16.9 (±9.37), compared to 5.9 (±4.30) for healthy controls. The mean depression subscale score for recovered subjects, was 10 (±6.32), compared to 3.8 (±3.76) for healthy controls, and the mean mania subscale score was 6.9 (±5.85), compared to 2.1 (±1.69) for healthy controls. The highest score on the MADRS for a healthy control subject was 6.0 and the highest YMRS score was 4.0. A sample of 17 provides 0.80 power to detect an effect size of 1.0 using a two-tailed test with a p value of ≤0.05.

Internal consistency of the BISS

Chronbach's alpha was 0.93 for the BISS total scale, 0.92 for the depression subscale and 0.90 for the mania subscale, indicating a reliable single factor total score and equally reliable subscale scores. For test–retest reliability, two BISS interviews were completed within three days by separate raters. Intra-class coefficients of correlation were uniformly high for BISS total score (0.96), BISS depression subscale score (0.98) and BISS mania subscale score (0.95). We also conducted the test–retest without the four major depressed subjects; coefficients were similar at BISS total score 0.94, BISS depression subscale 0.97 and mania subscale 0.92. To assess reliability for each of the BISS items we examined concordance of severity ratings between the two raters. We sought to identify items with a spread greater than two points for more than 10% (3/21) of the test–retest ratings. No items met this criterion. For 13/44 items (30%), retest ratings were within one point on all comparisons.

Discussion

These results indicate good concurrent validity in measuring and categorizing scientifically standard behavioral manifestations of mania and depression. In part, these results are consistent with the commonality in items for depression and mania in the scales. The results also indicate that the items we have added to encompass the bipolar spectrum also correlate with bipolar mood states as we had hypothesized. Not only were items added new to any bipolar rating scales, but items not included in the commonly used MADRS and YMRS were included in this scale (e.g. reduced sex drive, fearfulness, affective lability, impulsivity, risky behavior, elevated evening energy, sharpened thinking). The items added or combined into one rating scale provide a more comprehensive platform for adequate symptomatic assessment of BDs.

In this study of reliability and validity of the BISS we utilized established measures of depression, mania and overall functioning, as well as comparisons to healthy controls. We examined a large sample of bipolar subjects who varied in terms of severity, phases of illness, ethnicity/race and level of education. The YMRS and MADRS are currently the most commonly employed rating scales to assess manic and depressive symptoms respectively in BD clinical trials. The BISS yielded consistently high correlations across domains, e.g. 0.77 to 0.94 for the BISS depression subscale and MADRS scores and 0.78 to 0.91 for the BISS mania subscale and YMRS scores across mood states. The BISS provides enhanced discriminant capability in part because all of the items are obtained in the same structured fashion. However, the BISS has more items than commonly used scales for assessment of psychiatric symptoms, such as the BPRS (24 items), or using the YMRS (10 items) and MADRS (10 items) together. The administration time for the BISS ranges from 20 to 45 minutes depending on symptom endorsements. The potential increase in time to administer may make the BISS less attractive for use in practice and research settings.

The total BISS score yielded correlations of 0.81 with the MADRS and 0.74 with the YMRS for all subjects, in contrast to the lower correlations reported with the other psychiatric rating instruments recently published for BD, the BDSS, which correlated at 0.25 for a mania measure and 0.21 for an established depression measure (Dennehy et al., 2004). Our results of BISS correlations to MADRS for depressed patients,

DOI: 10.1002/mpr

0.83, are good, as the HAM-D (Hamilton Depression Rating Scale) and MADRS reported correlations of 0.56 to 0.96 in a recent study (Khan et al., 2004).

These results are important in that the shorter, traditional YMRS and MADRS and now BDSS scales do not include specific items encompassing domains of anxiety, impulsivity and mood lability that are included in the BISS. Additionally, the BISS provides an aggregate bipolar severity score not possible with the separately structured YMRS and MADRS scales. The BISS also provides structured questions to conduct and acquire scores. *Post hoc* interview guides have been developed for the YMRS and MADRS, but those versions have not been studied psychometrically.

A second area of interest was to ascertain whether any of the BISS scale scores correlated with global functioning scores using the GAF. For the full sample, the BISS total, depression and mania subscales correlated significantly with the GAF. The depression subscale showed a significant relationship to global functioning when we limited consideration to the depressed sample. The IDS-C and QIDS-C depression scales have also shown a relationship to overall improvement, as clinically meaningful improvements were associated with improved general functioning on the SF-12 for a bipolar sample in various phases of illness (Trivedi et al., 2004).

The BISS mania subscale showed a stronger relationship to the GAF in the manic/hypomanic subset of patients. The BISS mania subscale was significantly related to the GAF in manic/hypomanic and recovered subjects, but not in depressed or mixed subjects. These results suggest that GAF scores provide useful indicators of functional status in bipolar patients who are depressed, manic/hypomanic or recovered, but not for those in mixed states. Accordingly, these results suggest that BISS and GAF scale scores could both be useful in setting threshold parameters to define recovered, and possibly subsyndromal clinical status criteria. The GAF scores utilized in these comparisons were those for the lowest level of function over the prior month, a substantially different time frame than the prior week for the BISS. This introduced additional variation in the GAF scores which may have reduced degree of correlation in these comparisons.

Mixed episode subject scores showed no relationships with GAF. This may be consequent to the inherent complexity of mixed states, wherein a subset of depressive and manic symptoms, but not the full spectrum seen with a manic or depressive state, characterizes this still poorly understood and difficult to treat form of the disorder (Singh and Bowden, 2005). Also, we chose to include depressive episodes plus concurrent hypomanic episodes for the mixed episodes. Although hypomanic mixed states are considered dysfunctional due to agitation, hyperactivity, irritability and similar manic symptoms combined with depression (Akiskal and Benazzi, 2005), it is also possible that mixed states conceal dysfunction on a global rating due to the beneficial effects of some hypomanic symptoms.

The BISS total and subscale scores demonstrated excellent discriminant capability to distinguish between bipolar mood states and recovered status, consistent with a prior BISS study using a different sample of 20 persons (Bowden et al., 2007). In other areas, the BISS was indistinguishable in predictable ways, such as mixed episodes overlapping conceptually with both mania and depression subscales. The BISS total and depression and mania subscales also distinguished for most subsyndromal and recovered comparisons, and recovered and healthy control subjects.

The inclusion of all major BD clinical states strengthens the confidence for applying the discriminant data to the diverse illness states. The BISS appears capable of discriminating between relatively fine gradations, e.g. subsyndromal/symptomatic and each of the full syndromal states, as well as between recovered subjects. This is an important capability, as recent evidence indicates that bipolar patients have subsyndromal symptoms more often than syndromal ones and attention to such states of illness is increasing (Frye et al., 2006). We plan to report results of sensitivity to change with treatment utilizing the BISS when a sufficiently large sample for analysis is achieved.

None of the 44 items of the BISS had unacceptable reliability across raters according to the exploratory criteria we applied. During the initial development of the BISS, strong efforts were made to formulate items so that each item assesses only one symptomatic dimension or concept, for purposes of improving reliability and interpretability of items (Bowden et al., 2007). The clinician administered interview has required questions for each item and additional questions which are to be asked if responses to initial probe question are positive. As reviewed by Bagby et al. (2004), this step is essential to optimizing reliability. In the initial phase of reliability testing, we focused on achieving excellent

DOI: 10.1002/mpr

inter-rater reliability, and revised several items from the preliminary version of the BISS for this finalized version to improve reliability, as well as aid in other desired psychometric properties (Bowden et al., 2007).

This study has several limitations. This was a convenience sample with unbalanced groups. More balanced groups would strengthen the power for detecting differences and increase generalizability. We cannot evaluate whether the BISS can validly distinguish other possible subgroups of interest. For example, few subjects had severe symptomatology, as consent and cooperation for the required assessments would not have been feasible.

Participants with diagnoses were typically taking psychotropic medication, which may selectively improve symptoms and have impact on BISS item ratings. Unlike our first phase of psychometric characterization, in which we had the benefit of multiple blinded ratings of each subject, here, we as have others, relied on single ratings of each subject for the BISS, the YMRS and MADRS. One of the two methods utilized to categorize mood state was symptom severity ratings, as opposed to using a DSM-IV criteria structure, which inquires explicitly about general level of impairment. This means that hypomania, along with other mood states, was determined in some cases by a rating scale. The BISS is inherently likely to correlate to a second rating scale. Thus it is important to note that some correlations reported may be higher than had DSM-IV criteria been the only method used to establish mood states. Using one method for determining mood state would have strengthened the interpretation of these findings.

The discriminant validity assessed here was across mood state, the first time such a characterization has been made with a scale employed in BDs. However, we do not report here the discriminant characteristics of the BISS between bipolar depressed and major depressed subjects. The discriminant validity of the BISS reported here is limited to healthy controls. Strong evidence of symptom overlap and high rates of diagnostic comorbidity with BDs exists for impulsive and attentional control disorders as well as anxiety disorders (Kessler et al., 2005; Mantere et al., 2006). Studies of the similarities and differences across disorders with scales such as the BISS are warranted, as healthy control comparisons are not sufficient to comprise discriminant validity of a scale.

The sample of healthy controls is relatively small and is a convenience sample. Findings related to healthy controls are therefore preliminary and require replication. There is increasing evidence that use of healthy controls is relevant to better understand how a clinical population's characteristics compare to a general population. The BISS, or any change scale, would be less useful if incapable of distinguishing a healthy from an ill sample. Comparisons of healthy controls to subsyndromal bipolar persons are relevant both clinically and investigationally, as such scales are often utilized in patients and subjects in whom certainty of syndromal status is not established at the point of assessment.

The GAF is a widely used measure of overall functioning and disability (e.g. Parabiaghi et al., 2006; Ruggeri et al., 2000). However the GAF is limited for rating clinical status, and findings related to the GAF should be interpreted with caution. Further study with more comprehensive measures of functioning and quality of life could add utility to the BISS. A final limitation is the a priori structuring of depression and manic subscales and their items. The mania and depression subscale scores are significantly associated (data not shown). We attribute this to commonality of symptoms across mood states, such as anxiety or emotional lability. In a larger sample (unpublished data), the BISS comprised five distinct factors or behavioral domains, each with four or more items. Those domains captured the historical categories of depression and mania (not the same items comprising the two BISS subscales), plus irritability and psychosis, which are the four factors identified in most previous analyses of BD. A fifth domain, anxiety, was identified. Forty-one of the 44 items loaded onto one of these factors. Completion and publication of the factor analysis will determine the degree to which the depressive and manic subscale items conform with derived principal behavioral factors.

The BISS has established reliability and validity and a structured format with comprehensive characteristics that should allow its effective use in a wide range of studies on the psychopathology of BDs.

Declaration of Interests

The authors have no competing interests.

References

Akiskal HS, Benazzi F (2005). Toward a clinical delineation of dysphoric hypomania – operational and conceptual dilemmas. Bipolar Disord 7(5): 456–464.

Altshuler LL, Suppes T, Black DO, Nolen WA, Leverich G, Keck PE Jr, Frye MA, Kupka R, McElroy SL, Grunze H,

DOI: 10.1002/mpr

Kitchen CM, Post R (2006). Lower switch rate in depressed patients with bipolar II than bipolar I disorder treated adjunctively with second-generation antidepressants. Am J Psychiatry 163: 313–315.

Bagby RM, Ryder AG, Schuller DR, Marshall MB (2004). The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry 161: 2163–2177.

Benazzi F (2004). Depressive mixed state: a feature of the natural course of bipolar II (and major depressive) disorder? Psychopathology 37(5): 207–212.

Benazzi F, Koukopoulos A, Akiskal HS (2004). Toward a validation of a new definition of agitated depression as a bipolar mixed state (mixed depression). Eur Psychiatry 19: 85–90.

Berk M, Malhi GS, Mitchell PB, Cahill CM, Carman AC, Hadzi-Pavlovic D, Hawkins MT, Tohen M (2004). Scale matters: the need for a Bipolar Depression Rating Scale (BDRS). Acta Psychiatr Scand 110(Suppl): 39–45.

Bowden CL, Singh V, Thompson P, Gonzalez JM, Katz MM, Dahl M, Prihoda T, Chang X (2007). Development of the Bipolar Inventory of Symptoms Scale. Acta Psychiatr Scand 116: 189–194.

Cassidy F, Forest K, Murry E, Carroll BJ (1998). A factor analysis of the signs and symptoms of mania. Arch Gen Psychiatry 55: 27–32.

Cohen J (1988). Statistical Power Analysis for the Behavioral Sciences, 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dennehy EB, Suppes T, Crismon ML, Toprac M, Carmody TJ, Rush AJ (2004). Development of the Brief Bipolar Disorder Symptom Scale for patients with bipolar disorder. Psychiatry Res 127: 137–145.

Endicott J, Spitzer R (1978). A diagnostic interview: the Schedule of Affective Disorders and Schizophrenia. Arch Gen Psychiatry 35: 837–844.

Frye MA, Yatham LN, Calabrese JR, Bowden CL, Ketter TA, Suppes T, Adams BE, Thompson TR (2006). Incidence and time course of subsyndromal symptoms in patients with bipolar I disorder: an evaluation of 2 placebo-controlled maintenance trials. J Clin Psychiatry 67: 1721–1728.

Frye MA, Grunze H, Suppes T, McElroy SL, Keck PE Jr, Walden J, Leverich GS, Altshuler LL, Nakelsky S, Hwang S, Mintz J, Post RM (2007). A placebo-controlled evaluation of adjunctive modafinil in the treatment of bipolar depression. Am J Psychiatry 164: 1242–1249.

Kasper S, Lemming OM, de Swart H (2006). Escitalopram in the long-term treatment of major depressive disorder in elderly patients. Neuropsychobiology 54: 152–159.

Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 62: 617–627.

Khan A, Brodhead AE, Kolts RL (2004). Relative sensitivity of the Montgomery-Asberg depression rating scale, the Hamilton depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials: a replication analysis. Int Clin Psychopharmacol 19: 157–160.

Loo CK, Mitchell PB, McFarquhar TF, Malhi GS, Sachdev PS (2007). A sham-controlled trial of the efficacy and safety of twice-daily rTMS in major depression. Psychol Med 37(3): 341–349.

Mantere O, Melartin TK, Suominen K, Rytsala HJ, Valtonen HM, Arvilommi P, Leppamaki S, Isometsa ET (2006). Differences in Axis I and II comorbidity between bipolar I and II disorders and major depressive disorder. J Clin Psychiatry 67: 584–593.

Parabiaghi A, Bonetto C, Ruggeri M, Lasalvia A, Leese M (2006). Severe and persistent mental illness: a useful definition for prioritizing community-based mental health service interventions. Soc Psychiatry Psychiatr Epidemiol 41(6): 457–463.

Perlis RH, Miyahara S, Marangell LB, Wisniewski SR, Ostacher M, DelBello MP, Bowden CL, Sachs GS, Nierenberg AA (2004). Long-term implications of early onset in bipolar disorder: data from the first 1000 participants in the systematic treatment enhancement program for bipolar disorder (STEP-BD). Biol Psychiatry 55: 875–881.

Ruggeri M, Leese M, Thornicroft G, Bisoffi G, Tansella M (2000). Definition and prevalence of severe and persistent mental illness. Br J Psychiatry 177: 149–155.

Sachs GS (1990). Use of clonazepam for bipolar affective disorder. J Clin Psychiatry 51(Suppl.): 31–34, discussion 50–53.

Sachs GS, Thase ME, Otto MW, Bauer M, Miklowitz D, Wisniewski SR, Lavori P, Lebowitz B, Rudorfer M, Frank E, Nierenberg AA, Fava M, Bowden C, Ketter T, Marangell L, Calabrese J, Kupfer D, Rosenbaum JF (2003). Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). Biol Psychiatry 53: 1028–1042.

Schaffer A, Zuker P, Levitt A (2006). Randomized, double-blind pilot trial comparing lamotrigine versus citalopram for the treatment of bipolar depression. J Affect Disord 96: 95–99.

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 59(Suppl. 20): 22–33.

Singh V, Bowden CL (2005). Concepts surrounding the diagnosis and treatment of mixed states in bipolar disorders. Clin Approaches Bipolar Disorders 4: 35–43.

Thase ME, Macfadden W, Weisler RH, Chang W, Paulsson B, Khan A, Calabrese JR (2006). Efficacy of quetiapine monotherapy in bipolar I and II depression: a double-blind, placebo-controlled study (the BOLDER II study). J Clin Psychopharmacol 26: 600–609.

DOI: 10.1002/mpr

Tohen M, Bowden CL, Calabrese JR, Lin D, Forrester TD, Sachs GS, Koukopoulos A, Yatham L, Grunze H (2006). Influence of sub-syndromal symptoms after remission from manic or mixed episodes. Br J Psychiatry 189: 515–519.

Trivedi MH, Rush AJ, Ibrahim HM, Carmody TJ, Biggs MM, Suppes T, Crismon ML, Shores-Wilson K, Toprac MG, Dennehy EB, Witte B, Kashner TM (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. Psychol Med 34: 73–82.

Vatnaland T, Vatnaland J, Friis S, Opjordsmoen S (2007). Are GAF scores reliable in routine clinical use? Acta Psychiatr Scand 115: 326–330.

Young RC, Biggs JT, Ziegler VE, Meyer DA (1978). A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 133: 429–435.

Correspondence: Jodi M. Gonzalez, Department of Psychiatry, UT Health Science Center, San Antonio, Texas 78229, USA.
Telephone (+1) 210-567-5424
Fax (+1) 210-567-3759
Email: gonzalezjm1@uthscsa.edu