Original Article

Clinical decision-making using the General Behavior Inventory in juvenile bipolarity

Findling RL, Youngstrom EA, Danielson CK, DelPorto-Bedoya D, Papish-David R, Townsend L, Calabrese JR. Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. Bipolar Disord 2002: 4: 34–42. © Munksgaard, 2002

Objective: The General Behavior Inventory (GBI) is a questionnaire that has utility in the assessment of mood disorders in adults. The purpose of this study was to examine how the GBI might optimally be used in the assessment of youths.

Method: Children and adolescents between the ages of 5 and 17 years participated in this study. All youths were evaluated with the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS). Based on the K-SADS results, subjects were then assigned to one of four groups: a bipolar spectrum group, a depressive disorders group, a disruptive behaviors disorders group, and a no diagnosis group. Guardians completed a version of the GBI modified for parent reporting. Patients 10 years old or greater also completed the GBI as a self-report measure.

Results: There were 196 subjects who participated. Both parent report and youth self-report assigned patients to the appropriate diagnostic group with better than 74% accuracy. Combining information from multiple informants did not significantly improve diagnostic group assignment.

Conclusions: These data suggest that the GBI may be a useful adjunct in the diagnosis of mood disorders in youths, particularly when diagnostic specificity is more important than sensitivity.

**Robert L Findling,
Eric A Youngstrom,
Carla K Danielson,
Denise DelPorto-Bedoya,
Raïsa Papish-David,
Lisa Townsend and
Joseph R Calabrese**

Departments of Psychiatry and Pediatrics, University Hospitals of Cleveland/Case Western Reserve University School of Medicine, Department of Psychology, Case Western Reserve University, Cleveland, OH, USA

Corresponding author: Robert L Findling, MD, Director, Child and Adolescent Psychiatry, University Hospitals of Cleveland, 11100 Euclid Avenue, Cleveland, Ohio 44106-5080, USA. Fax: + 261 844 5883; e-mail: robert.findling@uhhs.com

Due to symptom overlap across psychiatric diagnoses, accurately determining whether or not a mood disorder is present or absent in a youngster may be a difficult undertaking. For example, disruptive behaviors are commonly seen in young people with a variety of psychiatric conditions. Youths suffering from major depression or bipolar disorder may have significant difficulties with aggression (1–4). In aggressive youngsters, the focus of clinical attention may be directed toward the symptoms of aggression. For this reason, the presence of a mood disorder may be missed in these patients. Conversely, young people with attention-deficit/hyperactivity disorder (ADHD) who are not suffering from a mood disorder might appear unhappy and have difficulties with attention. Therefore, these youngsters may be misdiagnosed as suffering from a depressive illness. What further complicates an accurate diagnosis of a mood disorder in a child or teenager is the fact that high rates of co-morbid disruptive behavior disorders have been reported in young people suffering from affective illnesses (2, 5–7).

For these reasons, a questionnaire that could be used to help delineate the presence or absence of a mood disorder in young people could be quite useful. The purpose of this study was to examine if an instrument that has been shown to have clinical utility in the assessments of adults with mood disorders, the General Behavior Inventory (GBI) (8), could make meaningful distinctions between young patients with mood disorders from other groups of children and adolescents.

Methods

Subjects

Eligible subjects were between the ages of 5 and 17 years (inclusive). Patients with a diagnosed or suspected pervasive developmental disorder (PDD), a psychiatric disorder due to a general medical condition, or evidence of mental retardation were not enrolled. Subjects were recruited for this protocol from an urban, university based division of child and adolescent psychiatry. Subjects were recruited from intake telephone calls placed to the division who expressed interest in participating in a research program. There are two research infrastructures within the division. One is a pediatric bipolar clinical research center (CRC). Patients seen within the auspices of the CRC are patients with bipolar spectrum disorder or youths who are at genetic high risk for developing the condition. In addition, the CRC actively recruits youths without psychiatric diagnoses to serve as research subjects. The other infrastructure is a pediatric psychopharmacology clinical trials infrastructure that recruits patients with a variety of different disorders.

The Institutional Review Board for Human Investigation of the University Hospitals of Cleveland approved the procedures of this protocol. Guardians of all subjects provided written informed consent for participation in this study. All youths provided assent.

Diagnosis

Diagnoses were made based on the results of a semi-structured diagnostic assessment. Both patients and their parent(s) were interviewed with either the Schedule for Affective Disorders and Schizophrenia for School-Age Children – Epidemiologic version (K-SADS-E) or the – Present and Lifetime version (K-SADS-PL) (9). The K-SADS assessment was administered by a child and adolescent psychiatrist or by one of 11 highly trained research assistants. Research assistants generally had at least 2 months of training prior to independently administering the K-SADS. Six of the research assistants had Master's degrees and five of the research assistants had Bachelor's degrees. Inter-rater reliability was established by having all research assistants trained to reach an overall kappa exceeding 0.85 over five interviews with an experienced rater prior to being able to independently interview youths and their families. Because many of these youths participated in pharmacological treatment studies, a child and adolescent psychiatrist confirmed the diagnoses based on the K-SADS interview in 48.5% of the patients. If

discrepant K-SADS diagnoses were reached, the physician and rater would have a consensus meeting in which final diagnoses were assigned. Although a specific screen for PDDs was not employed, non-verbal, poorly related youths were subsequently referred to study physicians in order to assess for the possibility of a PDD. The results of all K-SADS interviews were also reviewed at regularly scheduled research meetings by the principal investigator (RLF).

Based on the results of the diagnostic assessments, youths were assigned to one of four diagnostic groups. Youths meeting diagnostic symptom criteria for a diagnosis of bipolar I disorder, bipolar II disorder, cyclothymia, or bipolar disorder not otherwise specified were placed into the 'bipolar' group. Youths with bipolar spectrum disorders who also met diagnostic criteria for disruptive behavior disorders during identified periods of euthymia were also placed into this group. Patients assigned to the 'unipolar' group currently met diagnostic symptom criteria for a diagnosis of major depressive disorder and depressive disorder not otherwise specified. Patients in the unipolar group could also meet diagnostic symptom criteria for a co-morbid disruptive behavior disorder during periods of neutral mood. Those youths who did not suffer from a mood disorder and who had a diagnosis of ADHD, oppositional defiant disorder, or conduct disorder were included in the 'disruptive' group. Youths who did not meet diagnostic symptom criteria for any axis I disorder were placed into the 'no diagnosis' category.

Measure of interest

The GBI is an extensively validated questionnaire that is designed to assess for symptoms of depression and mania/hypomania (8). In adults, the GBI has been shown to have good discriminative validity and excellent reliability (10–13). There are also data to suggest that the GBI has good validity as a self-report measure in adolescents (14, 15). In addition, a slightly modified version of the GBI has shown promise as a parent-report measure of mood symptomatology in child and adolescent populations (16).

The GBI includes 73 Likert-type items reflective of mood symptomatology. Each item is rated on a four-point scale. When originally published, items on the GBI were given values of 1–4 and were scored on a 'case scoring' format. Symptoms that were rated as '1' or '2' were considered to be 'absent'. Items scored as '3' or '4' were considered to be 'present'. However, the authors of the GBI suggested that if the instrument was to be used in

a Likert-like fashion (e.g., each item could receive one of four scores), that items on the questionnaire should be scaled from 0 to 3 instead of from 1 to 4 (10). For this reason, it was decided to score the GBI based on the Likert scaling and not employ 'case scoring'. In this study, a score of '0' on a GBI item signified that the symptom in question was 'never or hardly ever' present. A score of '3' indicated that the symptom was 'often or almost constantly' manifest. Higher scores are indicative of greater degrees of psychopathology.

The primary caregiver of eligible youths were asked to complete the GBI as part of the young person's routine screening assessments. Adolescents aged 10–17 years were also asked to complete the GBI as a self-report measure. Ten years was chosen as the youngest age at which youth self-report on the GBI was obtained because it was felt that these were the youngest patients who could reliably understand the questions from the instrument. If there were concerns about a youth's reading ability, a member of the research team read items from the GBI to the teenagers.

The parent-completed version of the GBI was modified slightly so that the items on the GBI referred to the youth. Otherwise, the items on the GBI were unchanged and were kept in their original sequence. The same version of the parent-completed GBI was used for all subjects (children and adolescents). This was a conservative strategy, trying to preserve the original format of the validated adult version. However, response rates for individual items were examined due to concerns about the potential developmental appropriateness of item content across the broad age span of subjects. Parents typically responded to all items on the GBI. However, the most likely questions that were omitted by the guardians pertained to their child's/teenager's sexual functioning. Questionnaires with valid responses to at least 95% of the items were included in the statistical analyses. Typical items from the GBI and the parent-completed GBI are presented in Table 1.

Based on prior investigations, the items on the GBI were grouped into two scales (10, 14, 16). The Depressive Symptoms scale consists of 46 items with an alpha exceeding 0.97 when used as a parent-report measure. When utilized as an adolescent self-report measure, it has an alpha exceeding 0.96. The Hypomanic/Biphasic scale consists of 28 items (one of which is also included in the Depressive Symptoms scale). When used as a parent-report measure, the Hypomanic/Biphasic scale has an alpha of 0.96 (16). As an adolescent self-report measure, the Hypomanic/Biphasic scale has an alpha of 0.94 (14).

The GBI was administered to both the parent and the youth during the assessment process. Of note, the person performing the K-SADS interview did not have access to the results of the GBI prior to making a diagnosis.

Data analysis

Potential age effects were assessed at both the scale and item level. The correlations between age and scale or item scores were examined. A Bonferroni correction was used in conjunction with the item analyses to control type I error rate at 0.05 for 73 comparisons.

Correlations between parent and youth reports on the GBI were assessed using Pearson's correlation coefficients. Differences between means were compared with paired *t*-tests.

Logistic regression analysis (17) was employed in order to determine whether adding youth self-report data from the GBI reliably improved diagnostic decisions. Several regressions were run, collapsing the diagnostic groups into specific comparisons of clinical interest. The regression models began by predicting diagnostic groups using one scale from the parent-rated GBI. The parent Hypomanic/Biphasic scale was the first choice for discriminations involving bipolar spectrum disor-

Table 1. Sample items from the General Behavior Inventory

Original self-report version

a. Have you ever been sad, depressed, or irritable for several days or more without really understanding why?
b. Have people said you looked sad or lonely?
c. Have there been periods of several days or more when your friends or family told you that you seemed unusually happy or high – clearly different from your usual self or from a typical good mood?
d. Have there been times of several days or more when you did not feel the need for sleep and were able to stay awake and alert for much longer than usual because you were full of energy?

Modified parent-completed version

a. Have there been times of several days or more when your child was so sad that it was quite painful for him/her, or he/she felt that he/she couldn't stand it?
b. Has your child had periods when he/she was so down that he/she found it hard to start talking or that talking took too much energy?
c. Have there been times when your child had a strong urge to do something mischievous, destructive, risky, or shocking?
d. Has your child had times when his/her thoughts and ideas come so fast that he/she couldn't get them all out, or they came so quickly others complained that they couldn't keep up with your child's ideas?

ders. The parent-rated GBI Depression scale was the first choice for models classifying any mood disorder versus no diagnosis or versus a primary disruptive behavior diagnosis. These selections were made on conceptual grounds (e.g., it is the hypomanic and biphasic symptoms, and not the depressive symptoms, that would be expected to discriminate between bipolar versus unipolar spectrum disorders). Additionally, the choice of scales has been supported by prior research (14, 16). After controlling for the parent GBI variable, a forward stepwise entry procedure (based on conditional probability, with p < 0.05 as the entry criterion) tested whether either of the adolescent self-report GBI scales or the remaining parent scale made a significant contribution to discriminating between diagnostic groups.

These analyses are partially redundant with other analyses that are based on these data (14, 16). However, the previously reported work differed from this study in two important ways. First, the prior analyses considered adolescent and parent reports separately and did not attempt to combine the two. In addition, previous reports employed both the Depression and Hypomanic/Biphasic scale results in combination and not as separate scales.

In order to develop optimized cutting scores to fit the present sample, the following were calculated: sensitivity (i.e., the probability that a person with a mood disorder would score higher than threshold on the Depression scale, or that someone with a bipolar disorder would exceed threshold on the Hypomanic/Biphasic scale), specificity (i.e., the probability that someone without the disorder would score below the proposed threshold on the scale), positive predictive value (PPV; the probability that someone scoring higher than the GBI threshold actually has the disorder), and negative predictive value (NPV; the probability that someone scoring below the threshold truly does not have the disorder). Since investigators or clinicians, at times, may have different goals for using the GBI, two sets of cutting scores were developed based on the present sample. One was intended to achieve a sensitivity of at least 90%, and the other was intended to achieve a specificity of at least 90%.

In addition, receiver operating characteristic (ROC) curves were generated and areas under these curves (AUCs) were calculated in order to assess the ability of the GBI to discriminate between subject groups. ROC entails plotting the balance between the sensitivity and specificity of a diagnostic test while systematically moving the cut score across its full range of values.

Table 2. Primary diagnoses for the 190 youths who participated in this study

Diagnosis	n (%)
I. Bipolar group	70 (37)
A. Bipolar I disorder	30 (16)
B. Other bipolar disorders	40 (21)
II. Unipolar group	31 (16)
III. Disruptive behavior group	64 (34)
A. ADHD – combined type	46 (24)
B. ADHD – inattentive type	13 (7)
C. ADHD – hyperactive/impulsive type	1 (1)
D. Oppositional defiant disorder	2 (1)
E. Conduct disorder	4 (2)
IV. No diagnosis	25 (13)

The accuracy of a ROC can be quantified by calculating the AUC. Chance diagnostic performance corresponds to an AUC of 0.50, whereas an AUC of 1.0 indicates perfect diagnostic performance. The AUC is independent of cut score and does not assume that the underlying score distributions are normal. It is interpreted in terms of two children: one drawn randomly from the distribution of children with the target disorder and one selected randomly form the population of children without the problem. The AUC is the probability of the test correctly ranking the children in to their appropriate diagnostic groups. AUCs between 0.50 and 0.70 are characterized as showing low accuracy, 0.70–0.90 represent medium accuracy, and 0.90–1.00 denote high accuracy.

Averages are presented as means (standard deviations) unless otherwise noted. An analysis with the result of $p \leq 0.05$ was considered statistically significant. As this was not a confirmatory study, a Bonferroni correction was not employed.

Results

Subjects

There were 196 youths between the ages of 5 and 17 years with a mean age of 10.3 (3.1) years who participated in this study. Seventy percent (n = 138) of the subjects were males. Regarding race, 83.2% (n = 163) of the subjects were Caucasian, 11.7% (n = 23) of the subjects were African–American, and the rest of the patients were from other ethnicities. Further information regarding the study subjects can be found in Table 2. Six subjects had a primary diagnosis that did not fit into any of the diagnostic categories considered in this study. The primary diagnosis of these six youths included psychotic, anxiety, or adjustment disorders. The data for these six youths were not included in GBI

analyses. Fifty-five subjects had a GBI completed by both the subject and the subject's guardian. There were no differences on the GBI scales for either mean scores (all t-tests non-significant, p > 0.10) or variance (all Levene's F-tests non-significant, p > 0.05) when subjects diagnosed with the K-SADS-E were compared to those diagnosed with the K-SADS-PL.

As noted in other cohorts (see above), psychiatric co-morbidity was common in these young patients. Over half of the subjects (52.6%) met diagnostic symptom criteria for more than one axis I disorder. For patients with mood disorders, ADHD was the most common secondary diagnosis. Sixty percent of youths from the bipolar group (42/70) met diagnostic symptom criteria for ADHD during periods of euthymia. For patients in the unipolar group, 39% of youths (12/31) also met diagnostic criteria for ADHD. Four of the six patients with a primary diagnosis of oppositional defiant disorder or conduct disorder were also found to suffer from co-morbid ADHD.

Age differences

There were no significant associations (p < 0.05) between participant's age and their scores on the GBI scales. After Bonferroni correction, only one item had a statistically significant age effect. Parents reported more increased need for sleep in older adolescents (r = 0.30). Of note, the two items with sexual content were those most likely to be omitted, and they were more likely to be omitted in older subjects. In other words, parents of young children were more likely to report on these items whereas parents of teenagers were more likely to omit these items (although more than 90% of parents still answered them).

Parent–child agreement on mood symptoms

There were no significant differences on the GBI scores in terms of means [p > 0.05 for all t-tests, not significant (ns)] or standard deviations (p > 0.05 for all Levene's F-tests, ns) for the cases assessed with the K-SADS-E versus the K-SADS-PL. For the 55 adolescents with both self-report and parent-completed GBIs, parents and teenagers showed significant agreement about both depressive symptoms (r = 0.44, p = 0.001) and hypomanic/biphasic symptoms (r = 0.41, p = 0.002). Parents and teenagers reported similar degrees of mood symptomatology. For depressive symptoms, the mean GBI score for the parents was 32.6 (27.3) and for the adolescents it was 33.2 (25.1). The averages of these scores were not significantly dif-

ferent from each other (t = 0.17, df = 54, ns). For hypomanic/biphasic GBI-reported symptoms, the mean parent score was 21.9 (20.0) and the average score from the adolescents was 21.0 (15.3). As with the depressive symptoms, these means were also not significantly different from each other (t = 0.33, df = 54, ns). Intra-class correlation coefficients for absolute agreement (18) were comparable to the Pearson correlations: for depressive symptoms, parents and youths agreed 0.46, and 0.41 for hypomanic/biphasic symptoms.

Is there value from combining data from multiple informants?

Results from the logistic regression analysis are presented in Table 3. For all of the six analyses, the chosen parent GBI scale classified youths significantly better than chance, with Nagelkerke R^2 ranging from 0.46 to 0.77 and overall classification rates ranging from 79 to 94%. Interestingly, the second parent GBI scale never contributed significant additional information.

The two adolescent self-report scales also did not generally provide significant improvements in prediction, with two exceptions. The Hypomanic/Biphasic scale significantly improved the discrimination between bipolar spectrum versus all other disorders, with a change in one subject's classification. In the other exception, the bipolar spectrum versus disruptive behavior disorder classification improved from 79 to 87% as a result of correction of three cases.

Comparisons of different cut scores

When evaluating whether or not any mood disorder was present, the AUC derived from ROC analysis on the parent report version of the GBI was 0.88 [standard error (SE) = 0.03, p < 0.0005]. On the youth self-report version of the GBI, the AUC was 0.82 (SE = 0.06, p > 0.0005). When examining if a bipolar disorder was present or absent, the AUC on the parent report GBI was 0.88 (SE = 0.02, p < 0.0005), and for youths reporting about themselves on the GBI, it was 0.82 (SE = 0.06, p < 0.0005).

Table 4 presents diagnostic efficiency of statistics for several different possible cutting scores. Data are presented separately for the parent- and youth-reported GBIs, as relatively little information was gained by adding youth-reported GBIs when parent reports are available. Youth report data are also provided for clinicians for those instances in which parent reports are not readily available.

Based on these data, classifying a youth with Hypomanic/Biphasic scores of 17 or higher on the parent GBI would detect 90% of those youths with a bipolar spectrum disorder, but would erroneously identify about a third of youths without bipolar spectrum disorders. Alternatively, setting the cutting score at 36 or higher on the parent GBI would correctly identify 90% of those youths without bipolar spectrum disorders while still accurately identifying about 60% of youths with bipolar spectrum disorders.

Discussion

This study represents the first attempt to evaluate the clinical utility of the GBI in children and adolescents. The results of this work suggest that the GBI might be a useful tool in the assessment of depression and bipolar spectrum disorders in young people.

Youths and their parents do not always agree upon whether certain psychiatric symptoms are present or absent for the young person in question (19–21). However, parent–subject agreement about depressive and hypomanic/biphasic symptoms was substantially higher on the GBI than was expected. The correlation coefficients (r) on both scales of the GBI were found to be greater than 0.4. Based on previously reported results (22), such correlations are expected to be significantly below this. Although there was clear variability across the two sources of information on the GBI, it is likely

that the good reliability and content validity of the GBI scales contributed to these above average levels of inter-respondent agreement. In addition, most behavior checklists include items from a variety of domains. It is possible that the high level of concordance that was observed may be a reflection of the focused content of the GBI.

Based on these data, it appears that parent report on the GBI is diagnostically informative. Parent report appears similarly efficient in adolescents as in younger children. This is important when assessing young children for whom issues of reading ability and cognitive development would preclude the use of a self-report questionnaire. Although adolescent self-report on the GBI possesses good diagnostic efficiency in its own right (14), the adolescent-completed GBI does not appear to offer much incremental value when a parent report is available.

When the GBI is employed as a self-report measure and a 'case scoring' method is implemented, it has been proposed (8) that cutting scores of $13+$ for hypomanic/biphasic symptoms and scores of $+22$ for depressive symptoms be used for adults. However, when these scoring methods were employed in this cohort, they did not provide optimal results (Table 5). The data from this current study suggest that Likert scoring and different cut scores should be utilized in children and adolescents.

The GBI scales were found to have both acceptable sensitivity and specificity when the cut scores

Table 3. Logistic regression analyses using Depression (Dep) and Hypomanic/Biphasic (H/B) scales from the General Behavior Inventory to make diagnostic distinctions based on adult and youth self-report (adolescents only)

Comparison	Model fit		Classification[a]			Regression weights		
	χ^2 (2 df)	R^2	Overall	Dx	Non-Dx	Parent	Youth	Constant
(a) Any mood (n = 29) versus no diagnosis (n = 7)	23.72***	0.77	94%	97%	86%	0.20* Dep	–	−2.30*
(b) Bipolar (n = 19) versus no diagnosis (n = 7)	18.67***	0.74	92%	95%	86%	0.20* H/B	–	−2.26[ns]
(c) Any mood (n = 29) versus no mood (n = 26)	30.11***	0.56	84%	83%	85%	0.09*** Dep	–	−2.38**
(d.1) Bipolar (n = 19) versus not bipolar (n = 36)	27.44***	0.54	84%	68%	92%	0.10*** H/B	–	−2.97***
(d.2) Adding adolescent self-report	35.41***	0.65	82%	68%	89%	0.09** H/B	0.08* H/B	−4.75***
(e) Bipolar (n = 19) versus unipolar (n = 10)	11.71**	0.46	83%	90%	70%	0.10* H/B	–	−1.82[ns]
(f.1) Bipolar (n = 19) versus disruptive (n = 19)	19.42***	0.53	79%	68%	90%	0.10** H/B	–	−2.36**
(f.2) Adding adolescent self-report	25.45***	0.65	87%	84%	90%	0.09* H/B	0.08* H/B	−4.07**

*p < 0.05, **p < 0.005, ***p < 0.0005, all two-tailed.
[a] Classification rates are presented for overall accuracy of classification as well as the accuracy for the target diagnosis and the accuracy for the comparison group. Dx = diagnosis.

Table 4. Comparison of different cutting scores in terms of diagnostic efficiency statistics using the General Behavior Inventory in children and adolescents

Respondent	n	AUC	Base rate	Strategy	Cut score	Sensitivity	Specificity	PPV*	NPV*
Any mood disorder versus all others using the Depression Scale									
Parent	190	0.88	53.2%	Sensitivity 0.9	16+	0.90	0.66	0.75	0.86
				Specificity 0.9*	33+	0.75	0.90	0.89	0.76
				Maximum efficiency	27+	0.81	0.84	0.85	0.80
Youth	55	0.82	52.7%	Sensitivity 0.9	18+	0.90	0.65	0.74	0.85
				Specificity 0.9	42+	0.52	0.89	0.83	0.62
				Maximum efficiency	23+	0.86	0.69	0.76	0.82
Bipolar versus all others using the Hypomanic/Biphasic Scale									
Parent	190	0.88	36.8%	Sensitivity 0.9*	17+	0.91	0.68	0.62	0.93
				Specificity 0.9*	36+	0.60	0.90	0.78	0.79
				Maximum efficiency	17+	0.91	0.68	0.62	0.93
Youth	55	0.82	34.6%	Sensitivity 0.9	13+	0.89	0.47	0.47	0.90
				Specificity 0.9	32+	0.63	0.92	0.80	0.83
				Maximum efficiency	32+	0.63	0.92	0.80	0.83

Note: Sensitivity is the proportion of cases presumed to have the target diagnosis (based on K-SADS) who are correctly classified by the cut score. Specificity is the percentage of people presumed not to have the target diagnosis based on the K-SADS who are correctly classified by the cut score as not having the diagnosis. PPV = Positive Predictive Value, NPV = Negative Predictive Value, AUC = Area Under the Curve.

that are proposed herein are used (Table 4). The Depression and Hypomanic/Biphasic subscales also showed good PPV and NPV. PPV and NPV are probably of greater practical value than sensitivity and specificity because they both reflect the probability of making a correct diagnostic decision based on the proposed cutting score. The cut scores with high sensitivity would be most appropriate for screening applications where more specific measures could then be used to confirm or refute a formal diagnosis. The highly specific cutting scores are likely to miss most cases with the disorder, but they will likely be effective in identifying samples that are most likely to be suffering from mood disorders.

There are two ways that these findings might eventually be applied to other cohorts in research and clinical settings. One would be to use the logistic regression weights to calculate the probability that a specific individual is presenting with a bipolar spectrum disorder based on their GBI scores. The alternative to using logistic regression weights is to pick a set of cutting scores in order to make classification decisions.

There are several key strengths associated with a logistic regression approach. First, logistic regression analysis makes the greatest use of available information for each individual. In addition, logistic regression analysis allows for greater flexibility about making interpretive decisions. Another possible strength of regression equations is that they do not need to be modified when trying to assess for the presence or absence of different types of disorders. Instead, levels of probability can be determined. However, there are limitations to the application of a logistic regression approach. First, some practitioners may not be familiar with the practical applications of logistic regression coefficients. In addition, when employing logistic regression weights, a computer or scientific calculator is necessary in order to estimate odds ratios and probability for specific cases.

Advantages of the cut score approach include its convenience of calculation and ease of interpretation. Different sets of cut scores are provided in this report so that they might be used to give relatively high specificity or sensitivity, as needed. Limitations of the cutting scores include similar issues of generalizability that also affect logistic regression. The appropriateness of the cutting scores that are suggested in this study depends on the similarity of subsequent cohorts to the sample used here. Some measures of diagnostic efficiency, such as the sensitivity and specificity, are theoretically independent of the base rate of diagnoses. However, the predictive value of different cutting

scores depends heavily on the base rate of the disorders in question. It is possible to estimate new values for PPV and NPV using the sensitivity, specificity, and base rate (23), but the calculations would negate much of the convenience in using a cut score approach. An additional problem to cut scores is that cut scores are static. When relying on published recommendations, it is not clear how to adjust cutting scores in order to address specific needs for greater diagnostic sensitivity or specificity. Finally, there is the counter-intuitive problem that most individuals scoring right at the diagnostic threshold will not actually have the disorder (24) unless the base rate of the disorder is exceptionally high.

There are several limitations to this work. First, as this is the first study that has examined the clinical utility of the GBI in the young, this research needs to be replicated. In addition, this study was conducted under the auspices of two clinical research programs. Therefore, the ability to generalize these results to other clinical settings remains to be studied. Furthermore, not all subjects were diagnosed by a physician/clinician. The logistic regression coefficients in particular would benefit from cross-validation, as coefficients typically perform better in the original sample than they do in new groups. In addition, this study used a relatively small sample size. Furthermore, this work did not include youths with primary diagnoses of anxiety or psychotic disorders. Therefore, the utility of the GBI in other patient populations needs to be studied. In addition, self-report data were collected for less than half of the cohort.

Another limitation is the length of the GBI. However, the relatively large number of items may contribute to its excellent psychometric properties. Further research should address the possibility of developing a comparatively valid, but shorter set of scales.

The GBI scales showed good PPV and NPV in this cohort. Whereas specificity and sensitivity are theoretically unaffected by base rates, both PPV and NPV are highly dependent on the base rate of the disorders in question (23). The base rates of mood disorders exceeded 50% in the present sample, and more than a third of the participants had a bipolar spectrum diagnosis. These rates are higher than would typically be observed in community based clinical samples of youths. Both PPV and NPV can be recalculated depending upon the base rates for different conditions within a specific setting. As the base rate decreases, the NPV increases rapidly. Smaller base rates, however, pose a greater problem for PPV. If fewer patients at a given site have a bipolar disorder, a larger percentage of 'false positives' may be identified with the GBI.

Conclusion

The results of this study suggest that the GBI could be a clinically useful instrument. It appears that the GBI may be best utilized when assessing for the presence or absence of mood disorders, including bipolar spectrum disorders. In general, the GBI appears to be more specific than sensitive across the range of scores, making it a more powerful tool for ruling bipolar spectrum disorders 'out' rather than screening for them. Thresholds low enough to detect most cases with bipolar disorders (e.g., sensitivity exceeding 0.90) may misidentify a large number of youths without bipolar disorder.

The results of this study should not be interpreted as suggesting that the GBI should be used in isolation as a diagnostic instrument. Instead, it appears that the GBI, whether completed by a parent or a youngster, can provide valuable information. Further study with the GBI will clarify how this instrument may be best employed. As the GBI is over 70 items long, an abbreviated version of the questionnaire that retains this instrument's excellent psychometric properties might eventually prove to be even more useful.

Table 5. Diagnostic efficiency statistics using the General Behavior Inventory in children and adolescents using cutting and case scoring methods of Depue (8)

Respondent	n	Cut score	Sensitivity	Specificity	PPV	NPV	AUC
Any mood disorder versus all others using Depression Scale							
Parent	190	22+	0.23	0.97	0.89	0.52	0.88
Youth	55	22+	0.21	1.00	1.00	0.53	0.78
Bipolar versus all others using Hypomanic/Biphasic Scale							
Parent	190	13+	0.51	0.92	0.78	0.76	0.86
Youth	55	13+	0.26	1.00	1.00	0.72	0.80

PPV = Positive Predictive Value, NPV = Negative Predictive Value, AUC = Area Under the Curve.

References

1. Biederman J, Faraone S, Mick E et al. Child Behavior Checklist findings further support comorbidity between ADHD and major depression in a referred sample. J Am Acad Child Adolesc Psychiatry 1996; 35: 734–742.

2. Carlson GA. Mania and ADHD: comorbidity or confusion. J Affect Disord 1998; 51: 177–187.

3. Geller B, Warner K, Williams M et al. Prepubertal and young adolescent bipolarity versus ADHD: assessment and validity using the WASH-U-KSADS, CBCL, and TRF. J Affect Disord 1998; 51: 93–100.

4. Weiss B, Catron T. Specificity of the comorbidity of aggression and depression in children. J Abnormal Child Psychol 1994; 22: 389–401.

5. Angold A, Costello EJ. Depressive comorbidity in children and adolescents: empirical, theoretical, and methodological issues. Am J Psychiatry 1993; 150: 1779–1791.

6. Goodyer IM, Herbert J, Secher SM et al. Short-term outcome of major depression: I. Comorbidity and severity at presentation as predictors of persistent disorder. J Am Acad Child Adolesc Psychiatry 1997; 36: 179–187.

7. Rey JM. Comorbidity between disruptive disorders and depression in referred adolescents. Aust NZ J Psychiatry 1994; 28: 106–113.

8. Depue RA. General Behavior Inventory (Manual). Assessment Manual. Minneapolis, MN: University of Minnesota, 1987.

9. Kaufman J, Birmaher B, Brent D et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 1997; 36: 980–988.

10. Depue RA, Slater JF, Wolfstetter-Kausch H et al. A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: a conceptual framework and five validation studies. J Abnormal Psychol 1981; 90: 381–437.

11. Depue RA, Krauss S, Spoont MR et al. General Behavior Inventory identification of unipolar and bipolar affective conditions in nonclinical university population. J Abnormal Psychol 1989; 98: 117–126.

12. Klein DN, Dickstein S, Taylor EB et al. Identifying chronic affective disorders in outpatients: validation of the General Behavior Inventory. J Consult Clin Psychol 1989; 57: 106–111.

13. Mallon JC, Klein DN, Bornstein RF et al. Discriminant validity of the General Behavior Inventory: an outpatient study. J Personal Assess 1986; 50: 568–577.

14. Danielson CK, Youngstrom E, Findling RL et al. Youth self-report of hypomanic and depressed symptoms discriminates diagnostic categories. Poster presented at the national meeting of the American Psychological Society, June 9, 2000, Miami, Florida.

15. Klein DN, Depue RA, Slater JF. Inventory identification of cyclothymia. IX. Validation in offspring of bipolar I patients. Arch Gen Psychiatry 1986; 43: 441–445.

16. Youngstrom E, Findling RL, Danielson CK et al. Discriminant validity of parent report of hypomanic and depressive symptoms. Psychol Assess 2001; 13: 267–276.

17. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley, 1989.

18. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996; 1: 30–46.

19. Andrews VC, Garrison CZ, Jackson KL et al. Mother-adolescent agreement on the symptoms and diagnoses of adolescent depression and conduct disorders. J Am Acad Child Adolesc Psychiatry 1993; 32: 731–738.

20. Jensen PS, Rubio-Stipec M, Canino G et al. Parent and child contributions to diagnosis of mental disorder: are both informants always necessary? J Am Acad Child Adolesc Psychiatry 1999; 38: 1569–1579.

21. Loeber R, Green SM, Lahey BB. Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. J Clin Child Psychol 1990; 19: 136–143.

22. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychol Bull 1987; 191: 213–232.

23. Ingelfinger JA, Mosteller F, Thibodeau LA et al. Biostatistics in Clinical Medicine. New York: McGraw-Hill, Inc, 1994.

24. Hummel TJ. The usefulness of tests in clinical decisions. In: Lichtenberg JW, Goodyear RK, eds. Scientist–Practitioner Perspectives on Test Interpretation. Boston, MA: Allyn and Bacon, 1999; 59–112.