Developing a 10-Item Mania Scale From the Parent General Behavior Inventory for Children and Adolescents

Eric A. Youngstrom, Ph.D.; Thomas W. Frazier, Ph.D.; Christine Demeter, M.A.; Joseph R. Calabrese, M.D.; and Robert L. Findling, M.D.

Objective: Bipolar disorder is being diagnosed and treated in children and adolescents at a rapidly increasing rate, despite the lack of validated instruments to help screen for the condition or differentiate it from more common disorders. The goal of the present study was to develop and validate a brief (10 item) instrument to assess mania in a large sample of outpatients presenting with a variety of different DSM-IV diagnoses, including frequent comorbid conditions.

Method: Parents presenting to a Midwestern academic outpatient medical center for psychiatric evaluation of their child completed the Parent General Behavior Inventory (P-GBI), a 73-item mood inventory that comprises a 46-item depressive symptom scale and a 28-item hypomanic/biphasic scale (1 item is used in both scales), as part of a screening assessment that included a semistructured psychiatric interview of both the parent and the child to determine the child's diagnoses. The study was conducted between the years 1999 and 2004.

Results: Six hundred thirty-seven youths received a diagnostic assessment with either the Epidemiologic or Present and Lifetime Version of the Schedule for Affective Disorders and Schizophrenia for School-Age Children. A 10-item form derived from the 73-item P-GBI had good reliability (α = .92), correlated (r = 0.95) with the 28-item scale, and showed significantly better discrimination of bipolar disorders (area under the receiving operating characteristic [AUROC] curve of 0.856 vs. 0.832 for the 28-item scale, $p < .005$), with good precision for estimation of individual scores for cases up to 2 standard deviations elevated on the latent trait. The 10-item scale also did well discriminating bipolar from unipolar (AUROC = 0.86) and bipolar from attention-deficit/hyperactivity disorder (AUROC = 0.82) cases.

Conclusions: Findings suggest that parents most notice elated mood, high energy, irritability, and rapid changes in mood and energy as the prominent features of juvenile bipolar disorder.

(J Clin Psychiatry 2008;69:831–839)

Received Oct. 8, 2007; accepted Jan. 29, 2008. *From the Department of Psychology, University of North Carolina at Chapel Hill (Dr. Youngstrom) and Center for Pediatric Behavioral Health, Cleveland Clinic Foundation (Dr. Frazier) and Departments of Psychiatry, Case Western Reserve University and University Hospitals Case Medical Center (Drs. Calabrese and Findling and Ms. Demeter), Cleveland, Ohio.*

This research was supported by a Bipolar Disorder Clinical Research Center Grant (principal investigator, R.F.) from the Stanley Medical Research Institute, Chevy Chase, Md., as well as National Institute of Mental Health grants R01 MH066647 (principal investigator, E.Y.) and R01 MH073967.

The authors thank all of the families that participated in this project.

Financial disclosure appears at the end of this article.

Corresponding author and reprints: Eric A. Youngstrom, Ph.D., Department of Psychology, University of North Carolina, Davie Hall, CB3270, Chapel Hill, NC 27599-3270 (e-mail: Eric.Youngstrom@unc.edu).

The diagnosis of bipolar disorder in children and adolescents remains a contentious diagnostic issue.[1–3] Untreated bipolar disorder is likely to follow a progressive course, with each mood episode increasing the likelihood for future recurrences that may be more severe and resistant to treatment.[4–6] There also is concern that bipolar disorder may frequently be misdiagnosed in youths as attention-deficit/hyperactivity disorder (ADHD), conduct disorder, oppositional defiant disorder, or unipolar depression.[7–9]

Unfortunately, the diagnosis of juvenile bipolar disorder is difficult to make accurately. Prior research has tested several checklists and questionnaires as potential screening tools, comparing the diagnostic efficiency of positive test results against Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS) diagnoses as a criterion.[10–15] Checklists and screening measures possess considerable appeal, as they involve markedly lower costs in terms of training and administration, and they have the potential to promote the earlier identification and appropriate treatment of individuals experiencing bipolar disorder. These questionnaires also have the potential to reduce false-positive diagnoses. The Parent General Behavior Inventory (P-GBI)[10] is one such promising instrument.

The P-GBI has been evaluated in a large cohort of families presenting to an outpatient psychiatry clinic where the primary caregiver used the 73-item instrument to describe the lifetime presentation of mood symptoms in

children and adolescents aged 5 to 17 years.[10,14] The P-GBI has demonstrated exceptionally high internal consistency reliability, strong discriminant validity, and good diagnostic efficiency. High scores on the hypomanic/biphasic scale are associated with a large increase in the likelihood of a bipolar diagnosis. Scores on the 28-item hypomanic/biphasic scale can be used in high base-rate settings to correctly classify 4 of 5 children as having bipolar disorder versus any other diagnosis or bipolar disorder versus ADHD (arguably the most difficult differential diagnosis). Both the hypomanic/biphasic and depression scales of the P-GBI have demonstrated sensitivity to treatment effects,[16] and the high reliability of both scales also suggests that the P-GBI could be useful as an outcome measure.

However, the P-GBI has several important shortcomings, including its length and item wording, which is long and involves subtle nuances of context and duration. These characteristics were intended to enhance the clinical validity of the items and ensure that they captured the target construct of mood disorder. However, the unintended consequences included an increased level of reading difficulty (Flesch-Kincaid estimate of 12th-grade reading level), a longer questionnaire (10 pages in 12-point Times font), and increased parent burden as a result.

The goal of the present study was to develop a new scale derived from the item pool of the P-GBI that is shorter in length yet would maximize its value as a diagnostic aid (i.e., preserve diagnostic efficiency) for assessing bipolar spectrum disorder.

METHOD

Participants

The Institutional Review Board for Human Investigation of the University Hospitals Case Medical Center, Cleveland, Ohio, approved the procedures of this protocol. Parents provided written informed consent and youths provided informed assent, which were documented with age-appropriate forms. The diagnostic interview and questionnaires were completed as part of a screening protocol to determine potential eligibility for ongoing clinical trials addressing a wide range of diagnostic issues. Youths with a psychiatric disorder due to a general medical condition, a pervasive developmental disorder, or evidence of mental retardation were excluded. Participants were youths presenting at a midwestern urban outpatient research clinic specializing in the treatment of mood disorders but also conducting research on ADHD, conduct disorder, early-onset schizophrenia, and other diagnoses.

Measures

The P-GBI[10] is an adaptation of a well-validated instrument designed to screen for mood disorder in adult populations. The P-GBI consists of 73 Likert-type items rated on a scale from 0 ("Never or Hardly Ever") to 3 ("Very Often or Almost Constantly"), with high scores indicating greater pathology. The P-GBI has 2 scales: depressive symptoms (46 items, $\alpha = .97$) and hypomanic/biphasic (mixed) symptoms (28 items [1 item is used in both scales], $\alpha = .94$ in both the present and previously published subsamples); the published scoring instructions place 1 item[17] on both scales.[18]

Diagnostic Criterion

Primary diagnoses of the children and adolescents were made using either the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Epidemiologic Version[19] or the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version.[20] The diagnostic assessment was performed by either a child and adolescent psychiatrist or highly trained research assistants (17 bachelors-level assistants and 5 masters-level assistants). Assistants were trained to criterion by having them conduct 5 K-SADS interviews along with an experienced rater. New raters needed to lead 5 K-SADS interviews with an experienced rater and earn an overall $\kappa > 0.85$ on each in order to graduate from training.[21] Every 10th interview was done by 2 raters to maintain acceptable interrater reliability ($\kappa > 0.85$). Because many subjects also participated in pharmacologic clinical trials, diagnoses generated by using these semistructured diagnostic instruments were often confirmed with a clinical assessment performed by a child and adolescent psychiatrist (63% of cases). The researcher performing the K-SADS interview did not have access to the parent's results on the P-GBI during the diagnostic process.

Procedure

Primary caregivers contacted an outpatient Division of Child and Adolescent Psychiatry of University Hospitals Case Medical Center, Cleveland, Ohio, in order to have their child evaluated for potential participation in 1 of more than a dozen clinical trials open to a range of diagnoses, including conduct disorder, ADHD, unipolar depression, bipolar disorder, and early-onset schizophrenia. In addition, the sample was enriched with children at risk of bipolar disorder by means of referrals from an adult mood disorder clinic. The primary caregiver completed the P-GBI as part of an initial screening assessment that also included a semistructured psychiatric interview of both the parent and the child to determine the child's diagnoses. The study was conducted between the years 1999 and 2004.

Statistical Analyses

Preliminary analyses examined the missing item-level data. Although the factor structure of the P-GBI has been

examined previously, prior analyses relied on item parcels for the analysis.[10] However, because the goal of the present study was to develop a short form based on item characteristics and because a sample of adequate size was now available, a new item-level factor analyses was performed. The exploratory factor analyses began with a principal components analysis using the 3 most accurate decision rules to determine the number of factors to retain: Glorfeld's adaptation[22] of Horn's parallel analysis (GHPA), Velicer's method[23] of minimum average partials (MAP), and Cattell's scree test. O'Connor's syntax[24] to perform GHPA and MAP in SPSS, version 11.5 (SPSS Inc., Chicago, Ill.) was used. For the GHPA, 1000 simulations were conducted and compared observed eigenvalues to the 99th percentile of the empirical distribution.

After determining which items loaded onto the hypomanic/biphasic factor, cases according to their K-SADS diagnoses were divided into 2 groups: "any bipolar disorder" (including bipolar I, bipolar II, cyclothymia, or bipolar not otherwise specified [NOS]) versus "no bipolar disorder" (which included all remaining cases, regardless of diagnosis, provided that there was no lifetime history of a manic, mixed, or hypomanic episode). The size of the group differences on each item was then examined using t tests, with Cohen's d quantifying the effect size. Items were ranked in descending order by Cohen's d to determine which items best discriminated between the bipolar versus nonbipolar cases.[25] Cohen's d is directly related to other measures of group discrimination such as the area under the receiving operating characteristic (AUROC) curve analyses. [26,27]

Once the best discriminating items were identified, a second set of factor analyses was conducted to confirm that the items loaded onto a single factor.[28] Cronbach α and corrected item-total correlations measured the internal consistency of the 10-item form, and the correlation between the 10-item score and the original 28-item hypomanic/biphasic scale was also calculated. The diagnostic efficiency of the 10-item versus the 28-item version was compared using the z test to compare dependent AUROCs.[29] Next, Samejima's graded response model,[28] a form of item response theory appropriate for use with Likert-type items, was used to evaluate characteristics of the items (e.g., item discrimination and difficulty) and the total scale.[30] All of these analyses were performed 3 times—first on a random draw of half the sample (N = 318), then in a replicated fashion on the remaining half of the sample (N = 319). Because the results were usually identical to 2 decimal places and never differed by more than .01, even for AUROC or reliability estimates, only the analyses from the combined sample are reported (additional details of the split-half analyses available upon request from the first author).

Whether the diagnostic efficiency of the test might change as a function of child gender or age or whether or not the criterion diagnosis had been independently confirmed by a psychiatrist was also tested. These analyses were performed by estimating the receiving operating characteristic (ROC) curves separately for each subsample and then comparing the AUROCS using the test of independent curves.[29] A logistic regression analysis tested whether these factors influenced accuracy in a multivariate manner, including potential interactions between age and short-form score or gender and short-form score.

In addition, the performance of the short form in discriminating bipolar spectrum versus ADHD and disruptive behavior disorder cases (arguably the most difficult differential diagnosis) and bipolar versus unipolar depression (also a complicated differential diagnosis, particularly in children) was examined. The performance of the short form was evaluated using Kraemer's quality-calibrated diagnostic efficiency statistics,[27] which consider the "level" of the test (i.e., the proportion of cases testing positive) when quantifying test performance.

Finally, the multilevel likelihood ratios for the short form were estimated.[31] These are based on the sensitivity and specificity of a test at different score ranges and make it possible to calculate predictive values for individual cases. Sensitivity, specificity, and the positive and negative predictive power were also estimated for the same score thresholds to facilitate comparison to other instruments.

Exact p values are reported so that readers can compare significance to conventional ($p < .05$, 2-tailed) or more conservative thresholds for statistical significance.

RESULTS

Subjects

Six hundred thirty-seven subjects were enrolled and completed this study. Sixty-one percent (N = 388) were male, and 79% (N = 506) were white; 14%, black; and 7%, of other ethnicity. Youths ranged in age from 5 to 17 years (mean [SD], 11.3 [3.3]). The youths lived with their biological mother in 90% of cases, with the parents still living together in 46% of the cases. Across annual family income levels, the middle 50% of the sample had annual family incomes in the range of $20,000 to $40,000. Overall, the majority of the sample would be characterized as middle class but with about 20% having low income and educational attainment and 25% having relatively higher levels of both income and education (more details below).

According to K-SADS results, 131 youths (21%) met criteria for a unipolar mood disorder, including major depressive disorder, depressive disorder NOS, or adjustment disorder with disturbance of mood; 178 (28%) met criteria for bipolar I (with 3 cases meeting criteria for schizoaffective disorder, bipolar type); 116 (18%) met criteria for bipolar II (N = 12), cyclothymia (N = 38), or bipolar NOS (N = 66); 130 (20%) met criteria for disruptive

behavior disorders and/or ADHD (91 ADHD-combined type, 18 ADHD-predominantly inattentive, 5 ADHD-hyperactive type, 7 oppositional defiant disorder without comorbid ADHD, 7 conduct disorder without comorbid ADHD, and 2 with ADHD not otherwise specified [1 of which had an age at onset > 7 years]); and 60 that did not meet criteria for any Axis I disorder assessed by the K-SADS. Nineteen participants had other diagnoses (including 2 with schizophrenia, 1 with schizoaffective disorder-depressive type, 1 with generalized anxiety disorder, 3 with obsessive-compulsive disorder, 5 with post-traumatic stress disorder, 2 with cannabis dependence, 4 with enuresis, and 1 with an adjustment disorder of unspecified type). Three participants had "rule outs" or "diagnoses deferred."

There are a variety of different operational definitions possible for bipolar NOS.[32] Almost all of those cases considered bipolar NOS in the present sample showed a sufficient number and severity of manic symptoms but without the 1-week duration required for a strict DSM-IV diagnosis of a manic or mixed episode.[33]

The diagnostic categories above are based on primary diagnoses, in which mood disorder was considered to take precedence over other comorbid diagnoses for the purposes of the present study; 61% of the participants met criteria for more than 1 Axis I diagnosis, and 5 participants had as many as 6 Axis I diagnoses. The median number of Axis I diagnoses was 2. ADHD was the most common secondary diagnosis, appearing in 201 (68%) of the 294 youths with bipolar spectrum disorders and 154 (45%) of the 343 cases without bipolar diagnoses. For the secondary analyses comparing bipolar spectrum versus ADHD, the presence of any ADHD diagnosis superseded other nonbipolar or ADHD diagnoses.

Preliminary Analyses

Missing data. The majority of parents (72%) answered all of the P-GBI items; 98% of the items were complete. The 3 most commonly omitted items were number 70, "Have there been times of several days or more when your child lost all sexual interest?" (omitted by 84 parents); item number 61, "Have there been periods of a couple days or more when your child's sexual feelings and thoughts were almost constant, and he/she couldn't think about anything else?" (omitted by 35 parents); and item number 33, "Has your child experienced times of several days or more when he/she felt as if he/she was moving in slow motion?" (omitted by 23 parents). Parents were more likely to omit the items with sexual content for older children ($t = 4.19$, $df = 631$, $p < .0005$, $d = .33$), and parents were slightly more likely to omit items with sexual content for girls than boys ($\chi^2 = 4.23$, $df = 1$, $p = .040$). Children with complete P-GBI item data tended to be slightly younger, $t = 2.58$, $df = 631$, $p = .010$, $d = .21$ (effect sizes of .20 are considered "small"). No other item was missing more than 3.5% of the time. There were no other demographic or diagnostic differences (i.e., rate of bipolar vs. nonbipolar diagnoses) evident between those with complete versus partial GBI data (all bivariate p values > .10).

Factor analyses. Exploratory factor analyses tested the adequacy of the 2-scale format widely used with the GBI. The scree plot indicated 3 factors, GHPA suggested the retention of 4 factors, and MAP suggested 6. Based on a promax rotation, the 3-factor solution was interpretable. Solutions with a larger number of factors did not change the interpretation of the first 3 factors, whereas the subsequent factors tended to have fewer than 4 indicators with sizeable loadings and were not readily interpretable. The 3-factor solution basically consisted of a depression factor (rotated eigenvalue of 21.58), a hypomanic/biphasic factor (eigenvalue of 17.40), and a set of endogenous and somatic depressive symptoms (eigenvalue of 19.60). Essentially, the factor structure replicated what has been described by Depue[18] and has been implicitly used by others in constructing scale scores for the GBI, with the exception that the depression scale breaks into 2 correlated factors ($r = 0.65$ for the factor scores). Consistent with prior usage of the GBI and P-GBI, the mixed or biphasic items loaded primarily on the hypomanic/biphasic factor, although some showed substantial (> 0.40) cross-loadings on the depression factors.

Item discrimination of cases with versus without bipolar disorder. Table 1 presents the best 15 items (out of all 73 items) in descending order of effect size for mean differences between the bipolar versus nonbipolar groups. It is noteworthy that the majority (12 of 15) of the items with the largest group differences come from the hypomanic/biphasic factor, even though the nonbipolar group includes a large number of individuals diagnosed with ADHD as well as unipolar depression.

Internal consistency and reliable change indices. Factor analyses indicated that the 10-item short form measured a single factor, according to GHPA, MAP, and scree plots. The 15-item version included 2 factors based on the scree plot and MAP (but 1 factor according to GHPA); however, the second factor was a relatively small depression factor consisting of the 3 depression items plus a few small cross-loadings from other items, and the promax-rotated factors correlated, $r = 0.69$. The 10-item version was selected for subsequent analyses because it provided a substantial savings in length while preserving exceptional internal consistency and also because it was unifactorial.

Table 2 presents the α, the standard error of the measure,[34] and the critical values needed to be 90% and 95% confident that change in P-GBI scores reflects true change.[35] To calculate the Reliable Change Index (RCI) proposed by Jacobson and Truax,[35] one would simply take the difference between the case's 2 P-GBI scores and

Table 1. Fifteen Items From the P-GBI That Best Discriminate Bipolar Spectrum Disorders Versus All Other Diagnoses

Rank[a]	Original Item No.	Scale[b]	Content	Cohen's d	Estimated AUROC[c]
1	53	Biphasic	Days or more depressed/irritable, then days or more extremely high, elated, overflowing with energy	1.19	0.80
2	54	Hypomanic	Unusually happy and intensely energetic, but everything gets on nerves and makes angry	1.16	0.79
3	19	Biphasic	Mood/energy shifts rapidly from happy to sad or high to low	1.12	0.79
4	40	Biphasic	Feelings/energy are generally up or down but rarely in the middle	1.12	0.79
5	4	Hypomanic	Days unusually happy and intensely energetic, yet also physically restless, shifting activities	1.11	0.78
6	22	Hypomanic	Days or more of extreme happiness or energy yet also anxious or tense	1.01	0.76
7	11	Hypomanic	Days or more when others tell parent that child seems unusually happy or high—clearly different self	0.97	0.75
8	64	Hypomanic	Times when thoughts/ideas come so fast child cannot get them all out, or others complain they cannot keep up	0.95	0.75
9	27	Hypomanic	Days or more unusually happy and energetic yet also struggles with rage or urge to smash/destroy	0.95	0.75
10	31	Hypomanic	Days or more of extreme happiness and energy, and it takes over an hour to get to sleep at night	0.90	0.74
11	29	Depression	Days or more down and depressed and also physically restless, unable to sit still, shifting activities	0.85	0.73
12	58	Depression	Sad/depressed for days or more, interrupted by an hour to a day of extreme happiness and energy	0.84	0.72
13	42	Hypomanic	Strong urge to do mischievous, destructive, risky, or shocking acts	0.83	0.72
14	39	Depression	Feels low and depressed, yet also struggles to control rage or urge to smash/destroy things	0.82	0.72
15	35	Biphasic	Experiences both pleasurable and painful emotions more intensely than others	0.77	0.71

[a]The first 10 items (with largest effect sizes) are the 10 that were retained in the final version of the measure.
[b]Scale assignments are based on Depue et al.[39] labels for the original General Behavior Inventory.
[c]Estimated AUROCs are based on converting Cohen's d in the development sample.
Abbreviations: AUROC = area under the receiver operating characteristic, P-GBI = Parent General Behavior Inventory.

Table 2. Reliability, Standard Errors, and Critical Values for P-GBI Scores[a]

Variable	α[b]	SE_m[c]	SE_d[d]	95% Confidence[e]	90% Confidence[e]
10-item form	.92	2.29	3.24	6.34	5.34
15-item form	.94	2.85	4.03	7.90	6.65
Full length (28 item)	.94	4.32	6.10	11.97	10.07

[a]All values are based on Likert scoring (0 to 3 for each item).
[b]Cronbach α coefficient.
[c]Standard error of measurement.
[d]Standard error of the difference formulas presented in Pedhazur and Schmelkin.[34]
[e]P-GBI raw scores obtained by multiplying the SE_d by the appropriate normal curve deviate for 2-tailed distributions.
Abbreviation: P-GBI = Parent General Behavior Inventory.

divide it by the value for the standard error of the difference reported in Table 2. RCI scores greater than 1.65 would be considered 90% reliable, and RCI scores greater than 1.96 would be 95% likely to reflect real change and not just measurement error.

Samejima's graded response model[28] was used as implemented in MULTILOG 7.0.3[36] to evaluate both item characteristics and the overall reliability of the short form. The marginal reliability of the 10-item short form was estimated at 0.90, very similar to the α coefficient. The total test information was high for θ levels of –1.0 (corresponding to low levels of manic symptoms) through 2.3 (representing extremely high levels of manic symptoms), with an estimated standard error of the measure of 0.20 or smaller within this range. Levels of the manic "trait" were estimated for all 637 cases using full information maximum likelihood scoring. These scores correlated (r = 0.986) with the 10-item raw score. The weakest area of the test's performance was outside of the range in which it would be used clinically (i.e., the range involving very low levels of mania, corresponding to comparisons between individuals without any diagnosis vs. other individuals without bipolar disorder).

Correlations between short form and original form. The 10-item short form correlated (r = 0.95) with the original 28-item version. This is considered excellent preservation of content coverage. Even using stepwise regression, which capitalizes on chance structure in the data, produced an adjusted multiple r value of 0.975 based on a 10-item model—indicating that even choosing items solely on the statistical basis of maximizing correlation would not much improve the degree of content coverage.[37]

Diagnostic efficiency statistics. In ROC analyses discriminating bipolar versus nonbipolar youths, the original 28-item form achieved an AUROC of 0.832 (SE = 0.016) and the 10-item short form earned an AUROC of 0.856 (SE = 0.015). The short form discriminated bipolar cases significantly better than the full-length scale, z = 2.85, p < .005. Although it is likely that the performance of the short form will diminish somewhat when applied to a new

Table 3. Mean Scores on P-GBI 10-Item Mania Short Form Presented by Diagnostic Categories

		10-Item Raw Total		Estimated θ[a]	
Group	N	Mean	SD	Mean	SD
Bipolar I	178	17.12[b]	6.86	0.29	0.88
Other bipolar	116	12.90[c]	6.66	–0.21	0.86
Unipolar mood	131	6.10[d]	5.25	–1.19	0.75
ADHD/disruptive behavior disorder	130	7.03[d]	6.03	–1.04	0.86
Residual psychiatric diagnoses	19	6.63[d]	7.80	–1.18	1.13
No Axis I diagnosis	60	1.28[e]	3.82	–2.04	0.60
Total[f]	637[g]	10.17	8.10	–0.65	1.12

[a]Estimated θ based on scoring of a 10-item graded response model.
[b–e]Indicate homogeneous subsets, according to Games-Howell post hoc tests, $p < .05$.
[f]Overall $F = 95.17$; $df = 5,631$; $p < .00005$.
[g]Three participants had "rule outs" or "diagnoses deferred."
Abbreviations: ADHD = attention-deficit/hyperactivity disorder, P-GBI = Parent General Behavior Inventory.

Table 4. Association Between Maternal Education and Performance of the P-GBI 10-Item Mania Short Form

Maternal Education Level	%	AUROC (95% CI)	Cronbach α
Less than high school	6.2	0.76 (0.61 to 0.92)	.93
High school or GED	24.8	0.85 (0.79 to 0.91)	.91
Some college, business, or trade school	33.8	0.86 (0.81 to 0.91)	.91
College or university graduate	10.7	0.92 (0.86 to 0.98)	.93
Some graduate school or more	15.0	0.92 (0.81 to 0.95)	.93
Mother-reported P-GBI unavailable	9.6	0.81 (0.71 to 0.92)	.92

Abbreviations: AUROC = area under the receiver operating characteristic, GED = General Education Development, P-GBI = Parent General Behavior Inventory.

sample, it is 95% likely that the AUROC would fall in the range of 0.83 to 0.89 when tested in similar clinical samples; and the fact that the short form significantly outperformed the original form in the development sample suggests that the short form is unlikely to sacrifice any significant amount of diagnostic efficiency. Table 3 presents the mean scores on the 10-item short form associated with different diagnostic groups.

When the sample was limited to cases with bipolar disorder (including those with comorbid ADHD) versus those with ADHD (and perhaps other comorbid conditions, barring only bipolar disorders), the AUROC was 0.82 (SE = 0.021) for the 10-item short form versus 0.78 (SE = 0.023) for the 28-item version, $z = 2.97$, $p = .003$. Similarly, the short form did better at discriminating bipolar from unipolar depressed cases: AUROC = 0.86 (SE = 0.020) for the short form versus 0.82 (SE = 0.022) for the 28-item version, $z = 2.76$, $p = .006$.

Potential moderators of diagnostic efficiency. The AUROCs were estimated separately for each subgroup (seen by a psychiatrist, yes or no; male vs. female; ages 5 to 10 years vs. 11 to 18 years), and then a comparison of the 2 AUROCs was made using Hanley-McNeil test for independent curves. There were no significant differences in the AUROCs (all z values < 1.96, all p values > .05). The demographic factors that might have an effect jointly on classification accuracy were examined by including them as predictors in a logistic regression model predicting diagnostic status. Only the main effects for the short form and for having any ADHD diagnosis were statistically significant ($p < .05$), and the interaction terms indicated that there was no significant change in the performance of the short form due to gender, age, or ADHD status (all p values > .05). AUROCs were also calculated separately by stratifying on the mother's educational level (Table 4). Educational level is commonly used as a proxy measure of socioeconomic status, but, in this case, it offered an even more direct measure of whether the educational attainment of the respondent affected the reliability or validity of their responses on the P-GBI. Table 4 indicates that although there were no significant changes in reliability associated with level of education, the AUROC values tended to be lower for mothers with less education compared to those with more education ($p = .06$ for the comparison of "less than high school" vs. "college graduate").

Likelihood ratios. The likelihood ratios for 6 different ranges of test scores on the short form were calculated. Kraemer's quality-calibrated receiver operating characteristic (Q-ROC)[27] was used to identify statistically optimal places to divide the short form scores into segments. The best-calibrated sensitivity to bipolar diagnoses versus all others was achieved by treating raw scores of 1 or higher as a test positive, for example (sensitivity = 0.997, calibrated sensitivity = 0.970), and the best-calibrated specificity treated scores of 29 or higher as a positive test (specificity = 1.000, calibrated specificity = 1.000). The Q-ROC plot shows an unusual shape, indicating no clear winner in terms of maximal Cohen's κ value (Figure 1). Test scores in the range of 6 to 14 all produced κ coefficients of 0.50 to 0.56, which are clearly within sampling error of each other.

Estimated likelihood ratios were computed by dividing the distribution of scores on the short form into deciles and then collapsing deciles when the likelihood ratios were either similar or no longer increased monotonically. Table 5 provides the score ranges and their associated likelihood ratios. Because the 2 age groups did not show significantly different AUROCs in comparisons of either bipolar spectrum disorder versus all other disorders or bipolar spectrum disorder versus ADHD and because age in years was not a significant predictor in the logistic regression, parsimony dictated presenting a single table of likelihood ratios. Based on the Q-ROC results, scores of 0 were presented separately. More than 10% of the sample scored 0 on the short form, and the Q-ROC indicated that

Figure 1. Quality-Calibrated Receiver Operating Characteristic (Q-ROC) Curve for the 10-Item Mania Short Form Discriminating Bipolar Spectrum Versus All Other Diagnoses (N = 637)

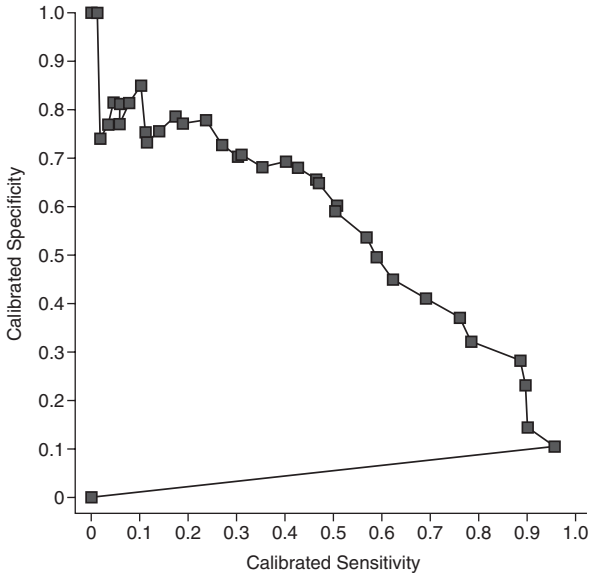

Table 5. Likelihood Ratios Associated With Test Scores on 10-Item Mania Short Form

Score Range[a,b]	Risk	Likelihood Ratio	Sensitivity, %	Specificity, %
0.0 to 0.9	Very low	0.01	100	0
1.0 to 4.9	Low	0.16	100	21
5.0 to 9.9	Low to neutral	0.56	94	54
10.0 to 14.9	Neutral	1.55	78	78
15.0 to 17.9	High	2.67	56	90
18.0 to 30.0	Very high	7.25	39	95

[a]If a respondent circled 2 options for a particular item, the average of the 2 was used as an item score. For example, if the parent circled 2 and 3 on an item, the item was scored as 2.5. Thus, it is possible to have noninteger scores on the short form.
[b]Sensitivity and specificity are reported for thresholds of 0, 1, 5, 10, 15, and 18.

treating scores of 1 or higher on the short form would be the optimal place to cut the test to maximize sensitivity (and negative predictive power). The result was 6 segments of test scores, with low scores (i.e., < 5) substantially decreasing the likelihood of a bipolar diagnosis and high scores (i.e., ≥ 18) increasing the likelihood of a bipolar diagnosis by a factor of more than 7.

DISCUSSION

The goal of the present study was to develop a brief mania scale from the P-GBI that parents could complete about their offspring as a screening device for juvenile bipolar disorder. Analyses indicated that it is possible to abbreviate the 28-item hypomanic/biphasic scale of the P-GBI into a 10-item form that possesses a slightly lower internal consistency but otherwise has psychometric properties that equal or exceed the characteristics of the full-length scale. The form consisted of the 10 items that maximally discriminated cases diagnosed with a bipolar spectrum disorder from cases with other nonbipolar diagnoses. The diagnostic efficiency of the 10-item form actually significantly exceeded the performance of the full-length scale in discriminating bipolar versus nonbipolar cases. It also is worth noting that the short form does well at discriminating bipolar disorder cases from ADHD cases, which, clinically, is perhaps the most difficult differential diagnosis.[7,8,38]

The 10-item form demonstrated a clear single factor structure based on a variety of criteria, whereas the factor structure of the 73-item P-GBI appears to be more complicated than initially thought.[10,39] Cases with bipolar I disorder scored significantly higher than all other diagnostic groups on the 10-item form. Cases with other bipolar spectrum diagnoses also scored higher than cases with ADHD/disruptive behavior disorders, unipolar mood, or other residual psychiatric diagnoses; and all groups with psychiatric diagnoses scored significantly higher than the "no Axis I diagnosis" group.

The 10-item form appears to have considerable potential as a screening device for juvenile bipolar disorder. Low scores are associated with very low likelihood ratios. These are likely to help "rule out" a bipolar diagnosis in most clinical settings, where the base rate of bipolar disorder is likely to be relatively low in the first place.[40,41] High scores also raise a clear "red flag": the likelihood ratios attached to extremely high screening scores are slightly higher than the increase in risk when a child has a bipolar parent[26] and are comparable to the likelihood attached to high scores on the full-length version of the P-GBI.[14] However, the likelihood ratios associated with high scores are not decisive by themselves. In most settings (all except those where the base rate of bipolar disorder exceeds 13%), the majority of children with a high score on the screening instrument will still not have a bipolar disorder (and even with 13% prevalence, the positive predictive value of scores of 18 or higher would still only be 51%). At the same time, such an elevated score offers a clear indication for a more detailed diagnostic evaluation. An attractive feature of the likelihood ratio approach is that it compels users to consider the base rate of bipolar disorder when interpreting the test result, thus avoiding some of the confusion and decision errors that can result from equating a positive test result with a diagnosis.[42]

The item content of the 10-item version is also clinically informative, as it was determined empirically by selecting items that most differentiated cases with bipolar diagnoses from other cases. This distinguishes it from instruments that are based directly on DSM criteria.[17]

Elated mood (described as "elated," "unusually happy," or "extreme happiness") was explicitly included in 8 of the 10 best discriminating items (see Table 1). This pattern strongly suggests that elated mood may be one of the "cardinal" symptoms that helps differentiate mania from other disruptive behavior disorders and ADHD in children.[43,44] Three of the top 10 items involved irritability, anger, or aggression, but always in the context of depressed (item 53) or elevated mood (items 54 and 27). This finding suggests that irritability may be associated with both polarities of mood (depressed and manic) but that mood still changes over time.[33] This result contrasts with the characterization of pediatric mania as primarily consisting of chronic irritable mood, perhaps even in the absence of other mood or energy changes.[45] Interestingly, grandiosity did not appear to be one of the better discriminators of bipolar disorder in youths, which is in contrast to other data.[43] Also prominent in the item content is an emphasis on changes in mood and energy, with mood states involving periods of "days or more" at each extreme (items 53, 19, 40), or mixed states involving a juxtaposition of elevated mood and irritability or anxiety (items 54, 4, 22, 27). Finally, it is noteworthy that none of the 46 depressive items on the P-GBI were among the best discriminators of bipolar disorder, in spite of clinical observations that bipolar depression may be associated more with atypical symptoms of depression (hypersomnia, increased appetite, rejection sensitivity, leaden paralysis).[46–48]

Limitations

Present findings must be qualified in several important ways. First, short forms can be developed with different goals in mind, and the same set of items will not perform equally well for different purposes. The present 10-item scale is intended to be a screening aid, and, to this end, maximizing its diagnostic efficiency was the primary concern. Other forms could be developed that potentially would have higher internal consistency, higher correlations with the full length scale, or greater sensitivity to treatment effects. Second, the operating characteristics of a new form need to be reevaluated in a new sample in which participants complete the scale in its proposed format as opposed to evaluating the performance of the items embedded in the original, full-length version.[49] New data are being collected using the 10-item scale with the items administered in the new format. Third, the mania scale needs to be validated in specifically those settings where use of the full-length version of the P-GBI is likely to be most problematic, i.e., settings where the base rate of bipolar disorder is low, the rate of other diagnoses is relatively high, and parent reading level is variable or low. The decrement in the performance of the mania scale in the groups with the lowest educational attainment needs further investigation and suggests that instruments with simpler reading levels may be needed for some settings. Finally, it will be important to evaluate the performance of the mania scale in more demographically diverse settings.

Conclusions and Future Directions

The 10-item mania scale represents a promising instrument. Even though its performance is likely to degrade somewhat when applied in new settings, it is likely to deliver diagnostic efficiency comparable to using the full-length version of the P-GBI but with considerable savings in terms of rater burden. It is interesting that even when focusing on youths obtaining high scores on a measure designed to be highly specific to bipolar disorder, there still will be many cases that would not meet "classic" criteria for bipolar disorder, despite showing marked emotional and behavioral disturbance. It would be valuable to use an instrument such as the present 10-item scale to identify a group of youths with elevated symptoms of mania and then follow them longitudinally to document the longitudinal evolution of their clinical presentation.

Such a sample would be most informative if the entry criteria focused on a moderately elevated score (such as a 12 or higher), because a moderately elevated level would likely capture a mix of cases both from the bipolar and nonbipolar spectrum.

Financial disclosure: **Dr. Calabrese** has received federal funding from the U.S. Department of Defense, Health Resources Services Administration, and the National Institute of Mental Health; has received research/grant support from Cleveland Foundation, NARSAD, Repligen, Stanley Medical Research Institute, Abbott, AstraZeneca, GlaxoSmithKline, Janssen, and Lilly; has served on advisory boards for Abbott, AstraZeneca, Bristol-Myers Squibb, France Foundation, GlaxoSmithKline, Janssen, Johnson & Johnson, Solvay/Wyeth; and has served as a CME faculty member for AstraZeneca, Bristol-Myers Squibb, France Foundation, GlaxoSmithKline, Janssen, Johnson & Johnson, and Solvay/Wyeth. **Dr. Findling** receives or has received research support, and has acted as a consultant and/or served on speakers bureaus for Abbott, AstraZeneca, Bristol-Myers Squibb, Cypress Biosciences, Forest, GlaxoSmithKline, Johnson & Johnson, Lilly, Neuropharm, New River, Novartis, Organon, Otsuka, Pfizer, Sanofi-Aventis, Sepracore, Shire, Solvay, Supernus Pharmaceuticals, and Wyeth. **Drs. Youngstrom** and **Frazier** and **Ms. Demeter** report no additional financial or other relationships relevant to the subject of this article.

REFERENCES

1. Nottelmann E, Jensen P. Current issues in childhood bipolarity. J Affect Disord 1998;51:77–80
2. McClellan J, Werry J. Practice parameters for the assessment and treatment of children and adolescents with schizophrenia. American Academy of Child and Adolescent Psychiatry. J Am Acad Child Adolesc Psychiatry 1997 Oct;36(suppl 10):177S–193S
3. Carlson G, Jensen P, Findling R, et al. Methodological issues and controversies in clinical trials with child and adolescent patients with bipolar disorder: report of a consensus conference. J Child Adolesc Psychopharmacol 2003;13:13–27
4. Geller B, Craney JL, Bolhofner K, et al. One-year recovery and relapse rates of children with a prepubertal and early adolescent bipolar disorder phenotype. Am J Psychiatry 2001;158:303–305
5. Post RM, Weiss SRB, Leverich GS. Recurrent affective disorder: roots in developmental neurobiology and illness progression based on changes in

gene expression. Develop Psychopathol 1994;6:781–813
6. Geller B, Tillman R, Craney JL, et al. Four-year prospective outcome and natural history of mania in children with a prepubertal and early adolescent bipolar disorder phenotype. Arch Gen Psychiatry 2004;61:459–467
7. Bowring MA, Kovacs M. Difficulties in diagnosing manic disorders among children and adolescents. J Am Acad Child Adolesc Psychiatry 1992 Jul;31(4):611–614
8. Biederman J, Klein RG, Pine DS, et al. Resolved: mania is mistaken for ADHD in prepubertal children. J Am Acad Child Adolesc Psychiatry 1998 Oct;37(10):1091–1096
9. Kim EY, Miklowitz DJ. Childhood mania, attention deficit hyperactivity disorder and conduct disorder: a critical review of diagnostic dilemmas. Bipolar Disord 2002;4:215–225
10. Youngstrom EA, Findling RL, Danielson CK, et al. Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. Psychol Assess 2001;13:267–276
11. Gracious BL, Youngstrom EA, Findling RL, et al. Discriminative validity of a parent version of the Young Mania Rating Scale. J Am Acad Child Adolesc Psychiatry 2002 Nov;41(11):1350–1359
12. Danielson CK, Youngstrom EA, Findling RL, et al. Discriminative validity of the general behavior inventory using youth report. J Abnorm Child Psychol 2003;31:29–39
13. Kahana SY, Youngstrom EA, Findling RL, et al. Employing parent, teacher, and youth self-report checklists in identifying pediatric bipolar spectrum disorders: an examination of diagnostic accuracy and clinical utility. J Child Adolesc Psychopharmacol 2003;13:471–488
14. Youngstrom EA, Findling RL, Calabrese JR, et al. Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. J Am Acad Child Adolesc Psychiatry 2004 Jul;43(7):847–858
15. Youngstrom EA, Gracious BL, Danielson CK, et al. Toward an integration of parent and clinician report on the Young Mania Rating Scale. J Affect Disord 2003;77:179–190
16. Youngstrom EA, Cooperberg M, Findling RL, et al. Identifying the most sensitive outcome measure for pediatric bipolar disorder. Presented at the 50th annual meeting of the American Academy of Child and Adolescent Psychiatry; October 14–19, 2003; Miami Beach, Fla
17. Hirschfeld RM, Williams JB, Spitzer RL, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. Am J Psychiatry 2000;157:1873–1875
18. Depue RA. General Behavior Inventory Assessment Manual. Minneapolis, Minn: University of Minnesota; 1987
19. Orvaschel H, Puig-Antich J, Chambers W, et al. Retrospective assessment of prepubertal major depression with the Kiddie-SADS-e. J Am Acad Child Psychiatry 1982;21:392–397
20. 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 Jul;36(7):980–988
21. Siegel S, Castellan NJJ. Nonparametric Statistics for the Behavioral Sciences. 2nd ed. Boston, Mass: McGraw-Hill; 1988
22. Glorfeld LW. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educ Psychol Meas 1995;55:377–393
23. Velicer WF. Determining the number of components from the matrix of partial correlations. Psychometrika 1976;41:321–327
24. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behav Res Methods Instrum Comput 2000;32:396–402
25. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988
26. Hodgins S, Faucher B, Zarac A, et al. Children of parents with bipolar disorder: a population at high risk for major affective disorders. Child Adolesc Psychiatr Clin N Am 2002 Jul;11(3):533–553
27. Kraemer HC. Evaluating Medical Tests: Objectives and Quantitative Guidelines. Newbury Park, Calif: Sage; 1992
28. Hambleton RK, Swaminathan H. Item Response Theory: Principles and Applications. Boston, Mass: Kluwer Nijofj; 1985
29. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–843
30. Embretson SE, Reise SP. Polytomous IRT models. In: Embretson SE, Reise SP, eds. Item Response Theory for Psychologists. Mahwah, NJ: Erlbaum; 2000:95–124
31. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994 Mar;271(9):703–707
32. Leibenluft E, Charney DS, Towbin KE, et al. Defining clinical phenotypes of juvenile mania. Am J Psychiatry 2003;160:430–437
33. Findling RL, Gracious BL, McNamara NK, et al. Rapid, continuous cycling and psychiatric co-morbidity in pediatric bipolar I disorder. Bipolar Disord 2001;3:202–210
34. Pedhazur EJ, Schmelkin LP. Measurement, Design, and Analysis: An Integrated Approach. Hillsdale, NJ: Erlbaum; 1991
35. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59:12–19
36. du Toit M, ed. IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT. Lincolnwood, Ill: Scientific Software, Inc; 2003
37. Cohen J, Cohen P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. 3rd ed. Hillsdale, NJ: Erlbaum; 1983
38. Geller B, Luby J. Child and adolescent bipolar disorder: a review of the past 10 years. J Am Acad Child Adolesc Psychiatry 1997 Sep;36(9):1168–1176
39. 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 Abnorm Psychol 1981;90: 381–437
40. Wozniak J, Biederman J, Kiely K, et al. Mania-like symptoms suggestive of childhood-onset bipolar disorder in clinically referred children. J Am Acad Child Adolesc Psychiatry 1995 Jul;34(7):867–876
41. Youngstrom EA, Findling RL, Youngstrom JK, et al. Towards an evidence-based assessment of pediatric bipolar disorder. Special issue: evidence-based assessment. J Clin Child Adolesc Psychol. In press
42. Hummel TJ. The usefulness of tests in clinical decisions. In: Lichtenberg JW, Goodyear RK, eds. Scientist-Practitioner Perspectives on Test Interpretation. Boston, Mass: Allyn and Bacon; 1999:59–112
43. Geller B, Williams M, Zimerman B, et al. Prepubertal and early adolescent bipolarity differentiate from ADHD by manic symptoms, grandiose delusions, ultra-rapid or ultradian cycling. J Affect Disord 1998;51:81–91
44. Carlson G. Bipolar disorder in children and adolescents: a critical review. In: Shaffer D, Waslick B, eds. The Many Faces of Depression in Children and Adolescents. Washington, DC: American Psychiatric Association; 2002:105–128
45. Biederman J, Faraone S, Mick E, et al. Attention-deficit hyperactivity disorder and juvenile mania: an overlooked comorbidity? J Am Acad Child Adolesc Psychiatry 1996 Aug;35(8):997–1008
46. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association; 2000
47. Goodwin FK, Jamison KR. Manic-Depressive Illness. New York, NY: Oxford University Press; 1990
48. Birmaher B, Ryan ND, Williamson DE, et al. Childhood and adolescent depression: a review of the past 10 years, pt 1. J Am Acad Child Adolesc Psychiatry 1996 Nov;35(11):1427–1439
49. Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. Psychol Assess 2000;12:102–111

Editor's Note: We encourage authors to submit papers for consideration as a part of our Focus on Childhood and Adolescent Mental Health section. Please contact Melissa P. DelBello, M.D., at mdelbello@psychiatrist.com

