Published in final edited form as:

J Clin Psychiatry. 2008 May ; 69(5): 831–839.

Developing a Ten Item Mania Scale from the Parent General Behavior Inventory for Children and Adolescents

Eric A. Youngstrom¹, Thomas W. Frazier², Christine Demeter^{3,4}, Joseph R. Calabrese^{3,4}, and Robert L. Findling^{3,4}

¹University of North Carolina at Chapel Hill ²Cleveland Clinic Foundation ³Case Western Reserve University ⁴University Hospitals of Cleveland

Abstract

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 in a large sample of outpatients presenting with a variety of different DSM-IV diagnoses, including frequent comorbid conditions.

Methods—Parents completed the Parent General Behavior Inventory (P-GBI), a 73 item mood inventory, as part of a screening assessment that included a semi-structured psychiatric interview of both the parent and the child to determine the child's diagnoses.

Results—637 youths completed the diagnostic assessment and the mood inventory. A 10 item form derived from this 73 item inventory had good reliability (.92), correlated .95 with the original scale, and showed significantly better discrimination of bipolar disorders (Area Under ROC Curve [AUROC] of .856 versus .832 for full scale, $p < .005$), with good precision for estimation of individual scores for cases up to two standard deviations elevated on the latent trait. The 10-item scale also did well discriminating bipolar from unipolar (AUROC = .86) and bipolar from ADHD cases (AUROC = .82).

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.

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 making it more likely that there will be 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 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-Epidemiological (KSADS) 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 version of the 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 has used the 73 item instrument to describe the lifetime presentation of mood symptoms in children 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 four out of five 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 that 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 (IRB) of the University Hospitals Case Medical Center approved the procedures of this protocol. Parents provided written informed consent and youths provided informed assent, 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 Parent General Behavior Inventory (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 two scales, a depressive symptoms (46 items, $\alpha = .97$) and a hypomanic/biphasic (mixed) symptoms (28 items, $\alpha = .94$ in both the present and previously published subsamples); the published scoring instructions place one 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-Epidemiological version (K-SADS-E)¹⁹ or the –Present and Lifetime version (K-SADS-PL).²⁰ The diagnostic assessment was performed by either a child and adolescent psychiatrist or highly trained research assistants ($n = 17$ bachelors level assistants and 5 masters-level assistants). Assistants were trained to criterion by having them conduct five K-SADS interviews along with an experienced rater. New raters needed to lead five K-SADS interviews with an experienced rater and earn an overall kappa $> .85$ on each in order to graduate from training.²¹ Every tenth interview was done by two raters to maintain acceptable inter-rater reliability (kappa $> .85$). Because many subjects also participated in pharmacological clinical trials, diagnoses generated using these semi-structured 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 GBI during the diagnostic process.

Procedure

Families contacted an outpatient Division of Child and Adolescent Psychiatry in order to have their child evaluated for potential participation in one 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 for 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 semi-structured psychiatric interview of both the parent and the child to determine the child's diagnoses.

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 three most accurate decision rules to determine the number of factors to retain: Glorfeld's adaptation of Horn's Parallel Analysis (GHPA)²² Velicer's method of Minimum Average Partial (MAP)²³ and Cattell's scree test. O'Connor's syntax²⁴ to perform GHPA and MAP in SPSS version 11.5 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 KSADS diagnoses were divided into two groups: "any bipolar disorder" (including bipolar I, bipolar II, cyclothymia, or bipolar 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). *T*-tests then examined the size of the group differences on each item, with Cohen's *d* quantifying the effect size. Items were ranked in descending order by *d* to determine which items best discriminated between the bipolar versus non-bipolar cases.²⁵ Cohen's *d* is directly related to other measures of group discrimination such as the Area Under the Curve²⁶ from Receiver Operating Characteristic (ROC) analyses.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's alpha 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, a form of IRT 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 three times – first on a random draw of half of the sample ($n = 318$), then replicated on the remaining half of the sample ($n = 319$). Because the results were usually identical to two 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 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's²⁷ quality calibrated diagnostic efficiency statistics, which consider the "level" of the test (i.e., the proportion of cases testing positive) when quantifying test performance.

Finally, the multi-level 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 ($p < .05$, two-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. Youth ranged in age from 5 to 17 years ($M = 11.3$, $SD = 3.3$). The youth lived with their biological mother in 90% of cases, with the parents still living together in 46% of the cases. 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 youth (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 for Bipolar I (with 3 cases meeting criteria for schizoaffective disorder, bipolar type); 116 (18%) for Bipolar II ($n = 12$), Cyclothymia ($n = 38$), or Bipolar NOS ($n = 66$); 129 (20%) 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, and 7 Conduct Disorder without

comorbid ADHD); and 60 that did not meet criteria for any Axis I disorder assessed by the KSADS. Nineteen participants had other diagnoses (including two with schizophrenia, one with schizoaffective disorder – depressive type, one with Generalized Anxiety Disorder, three with Obsessive/Compulsive Disorder; five with PTSD, two with cannabis dependence, four with enuresis, and one with an adjustment disorder of unspecified type).

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

These are primary diagnoses, where 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 one Axis I diagnosis, and five participants had as many as six Axis I diagnoses. The median number of Axis I diagnoses was two. ADHD was the most common secondary diagnosis, appearing in 201 of the 294 youths (68%) with bipolar spectrum disorders, and 154 of the 343 cases without bipolar diagnoses (45%). For the secondary analyses comparing bipolar spectrum versus ADHD, the presence of any ADHD diagnosis superseded other non-bipolar 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 three most commonly omitted items were #70, “Have there been times of several days or more when your child lost all sexual interest?” (omitted by 84 parents), item #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), and item #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: $t(631) = 4.19, p < .0005, d = .33$; and parents were slightly more likely to omit items with sexual content for girls than boys, $X^2(1 df) = 4.23, p = .040$. Children with complete P-GBI item data tended to be slightly younger, $t(631) = 2.58, p = .010, d = .21$ (where 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 ’s $> .10$).

Factor Analyses—Exploratory factor analyses tested the adequacy of the two-scale format widely used with the GBI. The scree plot indicated three factors, GHPA suggested the retention of four factors, and MAP suggested six. Based on a PROMAX rotation, the three factor solution was interpretable. Solutions with a larger number of factors did not change the interpretation of the first three factors, whereas the subsequent factors tended to have fewer than four indicators with sizeable loadings and were not readily interpretable. The three 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 and been implicitly used by others in constructing scale scores for the GBI, with the exception that the depression scale breaks into two correlated factors ($r = .65$ 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 ($> .40$) cross-loadings on the depression factors.

Item Discrimination of Cases with versus without Bipolar Disorder—Table 1 presents the best fifteen items (out of all seventy-three items) in descending order of effect size for mean differences between the bipolar versus nonbipolar groups. It is noteworthy that the majority (12 out 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 two factors based on the scree plot and MAP (but one factor according to GHPA); however, the second factor was a relatively small depression factor consisting of the three depression items plus a few small cross-loadings from other items, and the PROMAX rotated factors correlated $r = .69$. 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 alphas, 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 Index³⁶ proposed by Jacobson and Truax,³⁵ one would simply take the difference between the case's two P-GBI scores and divide it by the value for the standard error of the difference reported in Table 2. RCIs greater than 1.65 would be considered 90% reliable, and RCIs greater than 1.96 would be 95% likely to reflect real change and not just measurement error.

Samejima's Graded Response Model²⁸ was used as implemented in MULTILOG 7.0.3³⁷ to evaluate both item characteristics and the overall reliability of the short form. The marginal reliability of the ten-item short form was estimated at .90, very similar to the alpha coefficient. The total test information was high for theta 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 $r = .986$ with the ten-item raw score. The weakest area of the test's performance was outside of the range where it would be used clinically (i.e., the range involving very low levels of mania, corresponding to comparisons between individuals without any diagnosis versus other individuals without bipolar disorder).

Correlations between Short Form and Original Form—The ten item short form correlated $r = .95$ 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 .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.³⁸

Diagnostic Efficiency Statistics—In ROC analyses discriminating bipolar versus nonbipolar youths, the original 28 item form achieved an AUROC of .832 (standard error = .016), and the 10 item short form earned an AUROC of .856 (SE = .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 sample, it is 95% likely that the AUROC would fall in the range of .83 to .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.

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), then the AUROC was .82 (SE = .021) for the 10 item short form, versus .78 (SE = .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 = .85 (SE = .020) for the short form, versus .824 (SE = .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 versus female; ages 5 to 10 years versus 11 to 18 years), and then compared the two AUROCs using Hanley & McNeil's test for independent curves. There were no significant differences in the AUROCs, all z values < 1.96 , all $ps > .05$. 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 ($ps < .05$), 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 $ps > .05$). AUROCs were also calculated separately stratifying on the mother's educational level. Educational level is commonly used as a proxy measure of socioeconomic status, but in this case it offered an even more direct measure of the whether the educational attainment of the respondent affected the reliability or validity of their responses on the PGBI. Table 5 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" versus "college graduate").

Likelihood Ratios—The likelihood ratios for six different ranges of test scores on the short form were calculated. Kraemer's ²⁷ quality calibrated ROC (Q-ROC) 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 = .997, calibrated sensitivity = .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 kappa value. Test scores in the range of 6 to 14 all produced kappa coefficients of .50 to .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 3 provides the score ranges and their associated likelihood ratios. Because the two age-groups did not show significantly different AUROCs for either the bipolar spectrum versus all other or versus ADHD comparisons, 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 zero on the short form, and the Q-ROC indicated that 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 six 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 seven.

Discussion

The goal of the present study was to develop a brief mania scale from the General Behavior Inventory 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 ten 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 vs. nonbipolar cases. It also is worth noting that the short form does well at discriminating bipolar disorder from cases with ADHD, which clinically is perhaps the most difficult differential diagnosis.^{7, 8, 39}

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, 40} 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 depression, 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.^{41, 42} High scores also raise a clear “red flag”: The likelihood ratios attached to extremely high screening scores slightly higher than the increase in risk when a child has a bipolar parent²⁶ and 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.⁴³

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 eight of the ten 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.^{44, 45} Three of the top ten items involved irritability, anger, or aggression, but always in the context of depressed (item 53) or elevated mood (items 54 and 9). 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 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.³⁶ Interestingly, grandiosity did not appear to be one of the better discriminators of bipolar disorder in youths, in contrast to other data.⁴⁴ 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 forty-six 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 re-evaluated in a new sample where 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 lower, 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.

Acknowledgments

This research was supported by a Bipolar Disorder Clinical Research Center Grant from the Stanley Medical Research Institute as well as NIMH R01 MH066647 and NIMH R01 MH073967. We would like to thank Christine Demeter for all of her assistance with this project. We also thank all of the families that participated.

Dr. Findling receives or has received research support, acted as a consultant and/or served on a speaker's bureau 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.

References

1. Nottelmann E, Jensen P. Current issues in childhood bipolarity. *J Affect Disord* 1998;51:77–80. [PubMed: 10743840]
2. McClellan J, Werry J. American Academy of Child and Adolescent Psychiatry. Practice parameters for the assessment and treatment of children and adolescents with schizophrenia. *J Am Acad Child Adolesc Psychiatry* 1997;36(10 Suppl):177S–193S. [PubMed: 9432517]

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. [PubMed: 12804123]
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. [PubMed: 11156815]
5. Post RM, Weiss SRB, Leverich GS. Recurrent affective disorder: Roots in developmental neurobiology and illness progression based on changes in gene expression. *Development and Psychopathology* 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. [PubMed: 15123490]
7. Bowring MA, Kovacs M. Difficulties in diagnosing manic disorders among children and adolescents. *J Am Acad Child Adolesc Psychiatry* 1992;31:611–614. [PubMed: 1644722]
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;37:1091–1096. [PubMed: 9785721]
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. [PubMed: 12190710]
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. [PubMed: 11433802]
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;41:1350–1359. [PubMed: 12410078]
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. [PubMed: 12597697]
13. Kahana SY, Youngstrom EA, Findling RL, Calabrese JR. 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. [PubMed: 14977460]
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;43:847–858. [PubMed: 15213586]
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. [PubMed: 14607396]
16. Youngstrom, EA.; Cooperberg, M.; Findling, RL., et al. Identifying the Most Sensitive Outcome Measure for Pediatric Bipolar Disorder. Annual meeting of the American Academy of Child and Adolescent Psychiatry; October 2003; Miami Beach, FL.
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. [PubMed: 11058490]
18. Depue, RA. General Behavior Inventory Assessment Manual. Minneapolis, MN: 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. [PubMed: 7119313]
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;36:980–988. [PubMed: 9204677]
21. Siegel, S.; Castellan, NJJ. Nonparametric statistics for the behavioral sciences. Vol. 2nd ed.. Boston, MA: McGraw-Hill; 1988.
22. Glorfeld LW. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement* 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. *Behavior Research Methods, Instruments, & Computers* 2000;32:396–402.
25. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Vol. 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;11:533–553. [PubMed: 12222082]
27. Kraemer, HC. *Evaluating medical tests: objectives and quantitative guidelines*. Newbury Park, CA: Sage Publications; 1992.
28. Hambleton, RK.; Swaminathan, H. *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer Nijoff Publishing; 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. [PubMed: 6878708]
30. Embretson, SE.; Reise, SP. Polytomous IRT models. In: Embretson, SE.; Reise, SP., editors. *Item response theory for psychologists*. Mahwah, NJ: Erlbaum; 2000. p. 95-124.
31. Jaeschke R, Guyatt GH, Sackett DL. The Evidence-Based Medicine Working Group. 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? *JAMA* 1994;271:703–707. [PubMed: 8309035]
32. Leibenluft E, Charney DS, Towbin KE, et al. Defining clinical phenotypes of juvenile mania. *Am J Psychiatry* 2003;160:430–437. [PubMed: 12611821]
33. Findling RL, Gracious BL, McNamara NK, et al. Rapid, continuous cycling and psychiatric comorbidity in pediatric bipolar I disorder. *Bipolar Disord* 2001;3:202–210. [PubMed: 11552959]
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. [PubMed: 2002127]
36. 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;35:997–1008. [PubMed: 8755796]
37. du Toit, M., editor. *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software, Inc; 2003.
38. Cohen, J.; Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Vol. 3rd ed.. Hillsdale, NJ: Erlbaum; 1983.
39. Geller B, Luby J. Child and adolescent bipolar disorder: a review of the past 10 years. *J Am Acad Child Adolesc Psychiatry* 1997;36:1168–1176. [PubMed: 9291717]
40. 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. [PubMed: 7298991]
41. 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;34:867–876. [PubMed: 7649957]
42. 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
43. Hummel, TJ. The usefulness of tests in clinical decisions. In: Lichtenberg, JW.; Goodyear, RK., editors. *Scientist-practitioner perspectives on test interpretation*. Boston, MA: Allyn and Bacon; 1999. p. 59-112.
44. Geller B, Williams M, Zimmerman 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. [PubMed: 10743841]

45. Carlson, G. Bipolar disorder in children and adolescents: a critical review. In: Shaffer, D.; Waslick, B., editors. *The many faces of depression in children and adolescents*. Washington, D.C: American psychiatric Association; 2002. p. 105-128.
46. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders : DSM-IV-TR 4th ed, text revision*. Washington, D.C: American Psychiatric Association; 2001.
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. Part I. *J Am Acad Child Adolesc Psychiatry* 1996;35:1427–1439. [PubMed: 8936909]
49. Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. *Psychol Assess* 2000;12:102–111. [PubMed: 10752369]

Figure 1.
Quality-calibrated Receiver Operating Characteristic (Q-ROC) curve for the 10 item short form discriminating bipolar spectrum versus all other diagnoses ($N = 637$).

Table 1

Fifteen items that best discriminate bipolar spectrum versus all other diagnoses.

Rank	Original Item #	Scale ^a	Content	Cohen's <i>d</i>	Est. AUROC ^b
1	53	Biphasic	Days or more depressed/irritable, then days or more extremely high, elated, overflowing with energy	1.19	.80
2	54	Hypomanic	Unusually happy & intensely energetic, but everything got on nerves and made angry	1.16	.79
3	19	Biphasic	Mood/energy shifted rapidly from happy to sad or high to low	1.12	.79
4	40	Biphasic	Feelings/energy are generally up or down, but rarely in the middle	1.12	.79
5	4	Hypomanic	Days unusually happy & intensely energetic, yet also physically restless, shifting activities	1.11	.78
6	22	Hypomanic	Days or more extreme happiness or energy yet also anxious or tense	1.01	.76
7	11	Hypomanic	Days or more when others told you that child seemed unusually happy or high – clearly different self	.97	.75
8	64	Hypomanic	Times thoughts/ideas came so fast couldn't get them all out, or others complained they couldn't keep up	.95	.75
9	27	Hypomanic	Days or more unusually happy & energetic yet also struggled with rage or urge to smash/destroy	.95	.75
10	31	Hypomanic	Days or more of extreme happiness & energy, and it took over an hour to get to sleep at night	.90	.74
11	29	Depression	Days or more down and depressed, and also physically restless, unable to sit still, shifting activities	.85	.73
12	58	Depression	Sad/depressed for days or more, interrupted by an hour to a day of extremely happy & energetic	.84	.72
13	42	Hypomanic	Strong urge to do mischievous, destructive, risky, or shocking	.83	.72
14	39	Depression	Feeling low and depressed, yet also had to struggle to control rage or urge to smash/destroy things	.82	.72
15	35	Biphasic	Experiences both pleasurable and painful emotions more intensely than others	.77	.71

^a Scale assignments are based on Depue et al.'s 40 labels for the original GBI.

^b Estimated AUROCs are based on converting Cohen's *d* in the development sample.

Note: The first ten items (with largest effect sizes) are the ten that were retained in the final version of the measure.

Table 2

Reliability, standard errors, and critical values for Parent-Report GBI scores.

Variable	Alpha	SE _m	SE _d	95% Confidence	90% Confidence
Ten item form	.92	2.29	3.24	6.34	5.34
Fifteen item form	.94	2.85	4.03	7.90	6.65
Full length (28 item)	.94	4.32	6.10	11.97	10.07

Note: "Alpha" is Cronbach's alpha coefficient; "SE_m" is the standard error of measurement; "SE_d" is the standard error of the difference formulas presented in Pedhazur³⁴; "95% Confidence" and "90% Confidence" are P-GBI raw scores obtained by multiplying the SE_d by the appropriate normal curve deviate for two-tailed distributions. All values are based on Likert scoring (0 to 3 for each item).

Table 3
Mean scores on Parent GBI ten item “Mini-Mania” form presented by diagnostic categories.

Group	N	10 Item Raw Total		Estimated Theta	
		M	SD	M	SD
Bipolar I	178	17.12 ^a	6.86	+0.29	.88
Other Bipolar	116	12.90 ^b	6.66	-0.21	.86
Unipolar Mood	131	6.10 ^c	5.25	-1.19	.75
Disruptive Behaviors	133	7.03 ^c	6.03	-1.04	.86
Residual Diagnoses	19	6.63 ^c	7.80	-1.18	1.13
No Diagnosis	60	1.28 ^d	3.82	-2.04	.60
Total	637	10.17	8.10	-.65	1.12

Note: Superscripts indicate homogeneous subsets, according to Games-Howell post hoc tests, $ps < .05$. Overall $F(5, 631) = 95.17, p < .00005$. Estimated theta based on scoring of a 10 item graded response model.

Table 4

Likelihood ratios associated with test scores on 10 item hypomanic/biphasic short form

Score Range	Risk	Likelihood Ratio	Sensitivity	Specificity
0.0 to 0.9	Very low	.01	100%	0%
1.0 to 4.9	Low	.16	100%	21%
5.0 to 9.9	Low to Neutral	.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%

Note:

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

Table 5

Association between maternal education and performance of the PGBI mania short form

Maternal Education Level	Percentage of Sample	AUROC (95% C.I.)	Cronbach's Alpha
Less than High School	6.2%	.76 (.61 to .92)	.93
High School or GED	24.8%	.85 (.79 to .91)	.91
Some college, business, or trade school	33.8%	.86 (.81 to .91)	.91
College or University Graduate	10.7%	.92 (.86 to .98)	.93
Some Graduate School or More	15.0%	.92 (.81 to .95)	.93
Mother reported GBI unavailable	9.6%	.81 (.71 to .92)	.92