Comparing the Diagnostic Accuracy of Six Potential Screening Instruments for Bipolar Disorder in Youths Aged 5 to 17 Years

ERIC A. YOUNGSTROM, PH.D., ROBERT L. FINDLING, M.D., JOSEPH R. CALABRESE, M.D., BARBARA L. GRACIOUS, M.D., CHRISTINE DEMETER, B.A., DENISE DELPORTO BEDOYA, M.A., AND MEGAN PRICE, M.A.

ABSTRACT

Objective: To compare the diagnostic efficiency of six index tests as predictors of juvenile bipolar disorder in two large outpatient samples, aged 5 to 10 and 11 to 17 years, gathered from 1997 to 2002. **Method:** *DSM-IV* diagnosis was based on a semistructured diagnostic interview (Schedule for Affective Disorders and Schizophrenia for School-Age Children) with the parent and youth sequentially, blind to scores on the index tests. Participants were 318 youths aged 5 to 10 (50% with bipolar diagnoses) and 324 youths aged 11 to 17 (41% with bipolar diagnoses). Areas under the curve (AUCs) from receiver operating characteristic analyses and multilevel likelihood ratios quantified test performance. **Results:** Parent report (AUCs from 0.78 to 0.84 in both age groups) outperformed teacher (AUCs 0.57 in the younger sample and 0.70 in the older sample) or adolescent measures (AUCs 0.67 [General Behavior Inventory] and 0.71 [Youth Self-Report]) at identifying bipolar disorders. Combining tests did not produce clinically meaningful classification improvement. **Conclusions:** Parent report was more useful than teacher report or adolescent self-report on the index tests studied. Results generally replicated across both age groups. Parent report on these instruments could facilitate differential diagnosis of bipolar disorder in youths aged 5 to 17 years, especially by decreasing the rate of false-positive diagnoses. *J. Am. Acad. Child Adolesc. Psychiatry,* 2004;43(7):847–858. **Key Words:** bipolar disorder, sensitivity and specificity.

Although once considered rare, pediatric bipolar disorder may be more common than previously appreciated (Biederman et al., 2000; Hodgins et al., 2002). Young people have been given the diagnosis of bipolar disorder with increasing frequency (Carlson et al., 2003;

Naylor et al., 2002) For example, diagnoses of bipolar disorder have increased 260% in the period between 1994 and 2001 in the hospitalized wards for the Department of Child and Family Services for the State of Illinois (Naylor et al., 2002), and one market survey estimated that 95,000 youths were medicated for bipolar disorder in 2001 in the United States (Hellander, 2002). For this reason, whether many young people who are currently given the diagnosis of bipolar disorder truly suffer from the condition has come into question (Biederman et al., 1998; Pogge et al., 2001). Conversely, many youths who do suffer from bipolar disorder are often symptomatic for several years before a correct diagnosis is made (Findling et al., 2001; Geller and Luby, 1997).

An explanation for these diagnostic difficulties is that making an accurate diagnosis of bipolar disorder in the young may be quite challenging. Many of the

symptoms of bipolar disorder (such as hyperactivity, irritability, and aggressive behavior) occur in other psychiatric conditions that are common in the young, such as attention-deficit/hyperactivity disorder (ADHD), depression, and conduct disorder (American Psychiatric Association, 2001; Bowring and Kovacs, 1992; Carlson, 1998; Geller et al., 1998; Kim and Miklowitz, 2002). Young people with a primary diagnosis of bipolar disorder also often have comorbid conditions that can substantially complicate clinical presentation (Findling et al., 2001; Geller et al., 2000; Wozniak et al., 2002). Also, the longitudinal course of bipolar disorder may differ from the "classic" course of bipolar disorder described in adults: Many young people appear to have briefer episodes, fewer symptom-free periods, and higher rates of irritability (Findling et al., 2001; Geller et al., 2002a).

Unfortunately, pediatric bipolar disorder is a serious condition associated with chronic, substantial symptomatology and suffering, including increased mortality from suicide (Geller et al., 2002a). It is a common diagnosis among juvenile offenders (Pliszka et al., 2000), and manic symptoms appear to predict legal infractions across the life span (Hirschfeld et al., 2000). Although far less is known about evidence-based treatments for juvenile versus adult bipolar disorder, there are increasing data supporting interventions for acute mood episodes in youths (Frazier et al., 2001; Kafantaris, 1995; Kowatch et al., 2000; Wagner et al., 2002). For these reasons, accurate diagnosis of bipolar disorder could lead to improved outcomes for these psychiatrically affected youths.

The goal of this study was to compare six different screening instruments (Achenbach, 1991a–c; Depue et al., 1981; Gracious et al., 2002; Youngstrom et al., 2001) in terms of their diagnostic efficiency in facilitating the recognition of bipolar disorder in an adolescent sample (ages 11–17 years) and to cross-validate the performance of the parent and teacher measures in a younger sample (ages 5–10 years). A second goal was to compare the diagnostic value of parent, teacher, and youth report (Loeber et al., 1989, 1990) as well as determine the diagnostic information added by combining multiple informants and tests (Bird et al., 1992; Thuppal et al., 2002). The diagnostic performance of the "index tests" (Bossuyt et al., 2003) was evaluated in multiple ways, with the final goal of producing multilevel likelihood ratios to guide clinical interpretation of these tests in the most informative and practical manner possible (Jaeschke et al., 1994; Sackett et al., 2000). Multilevel likelihood ratios are a newer alternative to the concepts of sensitivity and specificity, and they extract more information from the test than would a single threshold score (Jaeschke et al., 1994; Sackett et al., 2000). The index tests have demonstrated medium or large differences in score distributions for bipolar and nonbipolar clinical comparison groups, suggesting that they might be valuable as tools for statistically aiding in the diagnostic classification of individuals (Biederman et al., 1995; Danielson et al., 2003; Findling et al., 2002; Gracious et al., 2002; Hazell et al., 1999; Lewinsohn et al., 2003; Youngstrom et al., 2001, 2003). These measures also possess substantial advantages as potential aids to differential diagnosis in a wide range of clinical settings: They are inexpensive and readily available and require minimal training to consistently administer and interpret (Drotar et al., 1995).

METHOD

Participants

The Institutional Review Board of Case Western Reserve University approved all procedures used in the study. Participants were recruited for more than a dozen different pharmacotherapy studies, depending on currently open protocols. Target diagnoses for protocols included bipolar disorder (bipolar I and cyclothymia or bipolar not otherwise specified [NOS]), unipolar depression, ADHD, conduct disorder, and aggressive behavior regardless of diagnosis. Recruitment was based on presenting symptoms and willingness to participate in treatment protocols. Advertisements and referrals described treatment studies, and those families interested in various treatment studies completed the diagnostic assessment as a screening or baseline evaluation. The sample was enriched by referrals of children whose parents had a diagnosed bipolar disorder and were participating in treatment or research at an affiliated adult mood disorders clinic. In addition, youths (including normal controls) were recruited by flyers and word of mouth to complete these descriptive psychometric instruments under the auspices of a Child/Adolescent Psychiatric Clinical Research Center.

Assessments took place at an outpatient clinic in an urban midwestern city. Inclusion criteria were (1) youths between 5 years 0 months and 17 years 11 months of age, (2) of either gender, (3) of any ethnicity, (4) presenting for an outpatient evaluation for which the youth provided written assent and the guardian provided written consent for participation, and (5) both the youth and the primary caregiver presented for the assessment. Exclusion criteria included the following: (1) ability of both the youth and the parent to communicate orally at a conversational level in English to complete the interview, (2) having a pervasive developmental disorder as determined by psychiatric history or psychiatric interview or having an Autism Screening Questionnaire score of 15 or higher (Beru-

ment et al., 1999), and (3) suspected moderate, severe, or profound mental retardation documented by educational history, standardized cognitive ability test scores of less than 70, or a Peabody Picture Vocabulary Test–Third Edition (Dunn and Dunn, 1997) score of less than 70 as a screener. All eligible participants completed the same assessment procedures, including the index tests and reference standard diagnostic interview, regardless of presenting symptoms or treatment study eligibility. The design was prospective: Data collection was planned before the index test and reference standard were performed (Bossuyt et al., 2003).

Measures

Reference Standard: Semistructured Diagnostic Interview Using the Schedule of Affective Disorders and Schizophrenia for Children. All participants and their families completed a semistructured diagnostic interview by a highly trained research assistant, using Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS) (Kaufman et al., 1997). K-SADS represents the most widely used semistructured diagnostic procedure used in investigations of juvenile bipolar disorder (Nottelmann et al., 2001). Diagnoses of bipolar I, bipolar II, cyclothymia, and bipolar NOS were made in strict accordance with *DSM-IV* diagnostic criteria (American Psychiatric Association, 1994). Failure to meet strict durational criteria was the most common reason for diagnosing bipolar NOS instead of one of the other bipolar diagnoses (cf. Leibenluft et al., 2003).

Research assistants ($N = 17$, ranging from a B.A. degree in psychology to a Ph.D. or M.D.) (see Findling et al., 2001 for detailed description of rater education and clinical experience) were trained to criterion by having them rate along while observing five K-SADS interviews by an experienced rater. New raters then led five K-SADS interviews with an experienced rater and achieved an overall κ of >0.85 at the symptom severity level and 1.0 agreement about the presence or absence of diagnoses on each to graduate from training. Acceptable interrater reliability (κ > 0.85 about symptom severity) was maintained by having joint rating sessions at every 10th interview or monthly, whichever happened first. The same interviewer worked with both informants, resolving discrepancies using best clinical judgment. As an additional reliability check, more than half of the participants enrolled in various treatment protocols and were independently diagnosed by a child and adolescent psychiatrist (60% in the younger sample; 65% in the older sample). The psychiatric diagnosis confirmed the K-SADS diagnoses in more than 95% of cases. Additionally, the joint reviews ensured that interrater agreement did not drop below κ of 0.85 at the item level.

Index Tests

Parent Young Mania Rating Scale. The Parent Young Mania Rating Scale (P-YMRS) (Gracious et al., 2002) is an 11-item questionnaire adapted from the Young Mania Rating Scale (YMRS) (Young et al., 1978). Parents rate their child's manic symptoms on five explicitly defined grades of severity, with item scores ranging from 0 to 4 (and three items ranging from 0–8). The P-YMRS yields a total score that can range from 0 to 56, with higher scores representing greater psychopathology. Ratings were based on the reported presence of symptoms over the past week. Internal consistency was adequate (α = .80 in the age 5–10 sample and .69 in the older sample).

General Behavior Inventory. The General Behavior Inventory (A-GBI) (Depue et al., 1989) is a 73-item self-report questionnaire measuring depressive, hypomanic, manic, and mixed ("biphasic") mood symptoms used with adolescents as young as age 11(Danielson et al., 2003; Findling et al., 2002). Respondents rate each symptom on a 0 (never or hardly ever) to 3 (very often or almost constantly) Likert-type scale, with higher scores indicating greater severity. The GBI yields two scale scores: a depressive (α = .96) and a hypomanic/biphasic score (α = .94) (Danielson et al., 2003). Current analyses used the hypomanic/biphasic score because preliminary findings indicate that this is the scale that best discriminates bipolar spectrum disorders from other diagnoses (Danielson et al., 2003; Findling et al., 2002). (Copies of the GBI are available from its author, Richard Depue [rad5@cornell.edu].)

Parent General Behavior Inventory. The Parent General Behavior Inventory (P-GBI) (Youngstrom et al., 2001) is an adaptation of the GBI, modified so that parents complete it to rate the depressive, hypomanic, manic, and biphasic mood symptoms of their children ages 5 to 17 years. The two scales of depressive and hypomanic/biphasic symptoms have strong construct validity (Youngstrom et al., 2001) and exceptionally high internal consistency (e.g., α of .97 for depression and .94 for hypomanic/biphasic in both age groups). The hypomanic/biphasic score has shown promise as a potential screener for bipolar disorder, based on preliminary analyses of a subsample of these youths (Findling et al., 2002; Youngstrom et al., 2001). (Copies of the P-GBI are available from the first author.)

Child Behavior Checklist. Parents also completed the 1991 version of the Child Behavior Checklist (CBCL) (Achenbach, 1991a), one of the most widely used instruments in research and clinical work (Sattler, 2002). The CBCL includes 118 problem behavior items rated from 0 (not at all typical of the child) to 2 (often typical of the child). The externalizing problems score has a 1-week test–retest stability coefficient of 0.93 and an α of .93 (Achenbach, 1991a).

Youth Self Report. Participating youths, ages 11 to 17, completed the Youth Self-Report Form (YSR) (Achenbach, 1991c). The YSR assesses the same behavior problems as does the CBCL, and most item content is identical. The externalizing score has a 1-week test–retest stability coefficient of 0.81 and an α of .89.

Teacher Report Form. The Achenbach Teacher Report Form (TRF) (Achenbach, 1991b) contains items and scales similar to those of the youth and parent report versions. The externalizing score on the TRF has a 15-day retest stability of 0.92 and an α of .90.

Procedure

The parent or guardian provided written consent for the participation of their child, and all youths provided written assent to participation. All participants and their families completed the intake assessment, which the involved the K-SADS diagnostic interview of the youth and parent by a highly trained research assistant. The research assistant also completed the YMRS (Young et al., 1978) and the Child Depression Rating Scale–Revised (Poznanski et al., 1984) as ratings of mood symptom severity and the Global Assessment of Functioning (*DSM-IV* Axis V) as a measure of overall impairment. The interviewer met with the adolescent first and then interviewed the parent(s) next, whereas the parent was interviewed first when younger children were assessed.

While the youth was being interviewed, parents completed the P-GBI, CBCL, and P-YMRS questionnaires. When the parent was

completing the K-SADS interview, then youths ages 11 to 17 were given the YSR and the GBI to complete. Youths and parents did not have access to each other's responses on the rating scales. The family identified one teacher who had the most extensive contact with the youth, and the family took a copy of the TRF and a stamped, addressed envelope and a fax number to give to the teacher to return the completed TRF. A reminder phone call often was made to families if the TRF was not returned in the next few weeks, but this was not systematically done. Families often brought back completed TRFs at a later visit. The K-SADS diagnoses were blind to the rating scales, which were scored after the completion of the interview.

Statistical Methods

To maximize the usable sample size and to avoid the possible introduction of bias due to missing data, we employed the multiple-imputation procedure developed by Graham and Schafer (Graham and Schafer, 1999; Schafer, 2002), imputing 10 sets of values separately for the younger and older samples.

Methods for Calculating and Comparing Diagnostic Accuracy

The primary criterion measure for all analyses grouped youths into two categories: (0) those with no diagnosis of a bipolar spectrum disorder, although multiple other Axis I diagnoses might be present and (1) those with any bipolar spectrum disorder (i.e., bipolar I, bipolar II, cyclothymia, or bipolar NOS) present regardless of whether other Axis I conditions might also be comorbid. Because previous investigations at this site found that no clinical scale from any of the Achenbach instruments either outperformed the externalizing scores or provided additional classification information above and beyond the externalizing scores, the current analyses examined only the externalizing scales (Kahana et al., 2003). The overall diagnostic efficiency of each test was quantified using nonparametric estimates of the area under the curve (AUC) from receiver operating characteristic analyses. We compared the diagnostic efficiency of the different index tests within each sample using the z test of dependent AUCs (Hanley and McNeil, 1983). We evaluated the reproducibility of results by using bootstrapping (500 resamples for each of the 10 imputed samples for both age groups, totaling 10,000 resamples) (Efron, 1982) and by comparing the generalizability of test performance across the two age groups by using the z test for independent AUCs (Hanley and McNeil, 1983).

Logistic regression analyses determined whether combinations of the index tests provided any incremental value after interpreting an individual index test (Hosmer and Lemeshow, 1989). Likelihood ratios (sensitivity divided by false alarm rate) quantified the diagnostic value of scores on each index test. To preserve more diagnostic information, likelihood ratios were estimated for multiple score ranges, dividing both samples into quintiles (Jaeschke et al., 1994; Sackett et al., 2000). Because scores on all the index tests were slightly positively skewed, we divided the highest scoring quintile into halves (i.e., 80th to 89th percentile and 90th percentile or higher) to examine whether extremely high scores would be useful in differentiating bipolar disorders. Likelihood ratios are the change in posterior odds associated with a particular test score. For example, a likelihood ratio of 3.0 would mean that the test result was associated with a tripling of the odds of a bipolar diagnosis.

RESULTS

Participants

A total of 642 youths, ages 5 to 17 years, participated, comprising a consecutive case series over the period from January 1996 to November 2002. Three fifths of the children came from intact families, with 20% living with divorced mothers, 5% with widowed mothers, and 6.5% with single mothers. Almost half of the caregivers earned less than $20,000 per year, and another third earned in the range of $20,000 to $40,000. For subsequent analyses, these were divided into two samples because youths needed to be at least 11 years old to complete the self-report measures. Table 1 presents demographic characteristics and K-SADS primary diagnoses. Consistent with epidemiological trends, the younger sample had a significantly higher rate of ADHD, and the older sample had a higher rate of unipolar depression. No adverse events were reported as a result of completing the index tests or K-SADS.

Missing Data Analyses

All eligible participants completed the reference standard (K-SADS). Index tests and mood symptom data were complete for 82.7% of the scores of interest in the younger sample and 81.4% of the older sample. The majority of the missing data were attributed to the TRF (45% overall return rate; smallest n = 125) and to the P-YMRS, which was added to the screening protocol in September 1999 (64% overall completion rate). Other measures ranged from 76.5% complete (A-GBI) to 95.3% complete (P-GBI). Except that children with missing TRFs tended to have slightly lower P-YMRS scores (r_{pb} = –0.27), the potential influence of missing data on observed scores was consistently negligible (next largest r_{pb} = 0.18). The data were well suited to the multiple-imputation procedure, both in terms of completeness and small and often nonsignificant patterns of missingness.

Diagnostic Efficiency Statistics

Table 2 presents descriptive statistics for the index tests and severity of mood symptoms separately for the criterion groups (K-SADS bipolar diagnosis present or absent). Table 3 presents correlations among the potential screening variables as well as the AUC from a

TABLE 1
Demographic and Diagnostic Characteristics

Characteristic	Ages 5–10 (N = 318)	Ages 11–17 (N = 324)
Age, years	8.3 (SD 1.6)	14.1 (SD 1.9)
Gender (% male)	210 (66)	185 (57)
Ethnicity, no. (%)		
African American	39 (12)	51 (16)
Hispanic	6 (2)	8 (2)
White	257 (81)	253 (78)
Other	16 (5)	14 (4)
Reference standard positive, no. (%)		
Bipolar I	97 (30.5)	79 (24.4)
Bipolar II, NOS, cyclothymia	63 (19.8)	53 (16.4)
Reference standard negative, no. (%)		
Unipolar depression (major depressive disorder or dysthymia)	33 (10.4)	103 (31.8)
ADHD or disruptive behavior without mood disorder	79 (24.8)	47 (14.5)
Residual (anxiety, posttraumatic stress disorder, psychotic disorders, or no Axis I)[a]	46 (14.5)	42 (12.9)
Any ADHD, no. (%)	211 (66.4)	147 (45.4)

Note: For present purposes, any mood diagnosis was considered "primary." Those with primary bipolar diagnoses also met criteria for 0 to 5 (median 2) other *DSM-IV* Axis I diagnoses. ADHD = attention-deficit/hyperactivity disorder.

[a] Two of the cases in the young sample and four cases in the adolescent sample were diagnosed with some variant of early-onset schizophrenia. These were included in the residual group because the sample size was insufficient to permit separate analysis, yet the bipolar versus schizophrenia differential diagnosis was considered too important to exclude the cases.

receiver operating characteristic analysis. For the older sample, the P-GBI performed significantly better than the CBCL externalizing, $z = 2.46$, $p < .05$. The P-GBI, CBCL, and P-YMRS earned AUCs significantly larger than for the A-GBI, YSR, or TRF in the older sample, z scores ranging from 2.03 (CBCL versus TRF) to 4.93 (all p values < .05). In the younger cohort, the three parent measures did not show significant differences in performance (all z scores < 0.70), but all three parent measures did substantially better than did the TRF diagnostically (z scores 5.38–5.90, $p < .0000005$). Table 4 presents the likelihood ratios associated with low, moderately low, intermediate, moderately high, high, and very high scores on each test.

Replication and Generalizability

The AUCs are generally comparable for the same measure across the two age groups (Table 3). Only the TRF externalizing score performed significantly different in the two groups, with teachers doing a better job of discriminating bipolar disorders in the 11- to 17-year-olds (AUC = 0.70) versus the younger cohort (AUC = 0.57, not significant), with $z = 2.38$, $p = .018$ (Hanley and McNeil, 1983). Bootstrapping revealed

minimal bias (largest bias was 0.00024 for AUC estimates).

Evaluating of Combinations of Index Tests

Logistic regression evaluated whether any of the other index tests significantly improved prediction after controlling for the most powerful index test. For the younger sample, the P-YMRS was the most powerful predictor. The CBCL externalizing score made a statistically significant contribution after controlling for the P-YMRS (Wald = 18.84, $p < .00005$). However, the combination of the two predictors improved classification accuracy only by less than 3%, suggesting that the interpretation of the two measures in combination was unlikely to be clinically meaningful. Neither the P-GBI nor the TRF externalizing scales made a statistically significant contribution after controlling for the P-YMRS. For the older sample, the P-GBI hypomanic/biphasic score entered first. The P-YMRS then still made a statistically significant contribution ($p < .0005$ for Wald test and χ^2 change), but the change in overall classification accuracy again was too small to be clinically meaningful (<3% improvement). None of the other index tests made a significant im-

TABLE 2
Index Test, Mood Symptom Severity, and Global Assessment of Functioning Distributions for Youths With and Without Bipolar K-SADS Diagnosis

Ages 5–10 (N = 318)	Nonbipolar (n = 158)		Bipolar (n = 160)		
Index Test	Mean	SD	Mean	SD	Cohen's d^a
P-YMRS	11.5	9.3	25.3	10.7	1.4
P-GBI	17.3	14.3	35.6	14.8	1.3
CBCL	60.0	13.1	73.9	7.4	1.4
TRF	58.5	12.5	61.4	12.7	0.2
YMRS	2.1	4.2	22.8	8.1	3.3
CDRS-R	27.9	17.5	30.9	12.8	0.2
GAF	60.4	16.2	49.7	8.1	–0.9

Ages 11–17 (N = 324)	Nonbipolar (n = 192)		Bipolar (n = 132)		
Index Test	Mean	SD	Mean	SD	Cohen's d^a
P-YMRS	10.7	7.7	20.6	9.0	1.2
P-GBI	15.2	13.1	34.7	15.7	1.4
A-GBI	19.6	13.7	29.2	16.4	0.7
CBCL	60.0	13.0	72.3	8.9	1.1
YSR	55.5	11.9	64.5	12.1	0.7
TRF	53.7	12.7	63.2	12.9	0.7
YMRS	0.7	2.4	22.6	8.5	4.5
CDRS-R	39.5	19.8	38.1	14.2	–0.1
GAF	56.6	16.5	50.6	8.7	–0.5

Note: P-YMRS = Parent Young Mania Rating Scale; P-GBT = Parent General Behavior Inventory; CBCL = Child Behavior Checklist; TRF = Teacher Report Form; YMRS = Young Mania Rating Scale; CDRS-R = Child Depression Rating Scale–Revised; GAF = Global Assessment of Functioning; A-GBI = Adolescent General Behavior Inventory; YSR = Youth Self-Report Form.

[a] Cohen's d of 0.2 constitutes a small effect size, 0.5 a medium, and 0.8 a large effect for the social sciences. All bipolar versus nonbipolar differences significant $p < .0005$, two tailed, unless otherwise indicated.

provement in prediction after controlling for the P-GBI and P-YMRS.

A final analysis compared the three informants on the Achenbach externalizing problems scale in the older sample. After controlling for parent report (CBCL externalizing), neither teacher nor youth report significantly improved the model ($p > .05$ for Wald tests and χ^2 change).

DISCUSSION

This is the first study to compare the diagnostic efficiency of six different rating scales as tools to facilitate the accurate diagnosis of bipolar disorders in youths of ages 11 to 17, and it is the first to compare the performance of parent and teacher measures in a younger sample of children ages 5 to 10 years. These results are consistent with previous findings pertaining to group differences (bipolar versus ADHD) on the Achenbach measures, where the largest mean differences have occurred on the CBCL, followed by the TRF, and then the YSR (Biederman et al., 1995; Carlson and Kelly, 1998; Carlson et al., 1998; Dienes et al., 2002; Geller et al., 1998; Hazell et al., 1999; Kahana et al., 2003). It is possible that teachers would provide more relevant information if completing other instruments besides the TRF, especially if the scales included more items assessing manic and hypomanic behaviors (Thuppal et al., 2002). The relatively good performance of parent report makes some sense, given the greater time that parents typically spend observing child behaviors, the cognitive developmental constraints on the reliability and validity of younger children's self-report (Sattler, 2002), and the documented

TABLE 3
Index Test Correlations and Global Measures of Diagnostic Efficiency

Correlations	P-YMRS	P-GBI	A-GBI	CBCL	YSR	TRF
Ages 5–10 (*N* = 318)						
P-YMRS	1.00	.71	—	.63	—	.19*
P-GBI	.71	1.00	—	.69	—	.24
CBCL	.63	.69	—	1.00	—	.33
TRF	.19*	.24	—	.33	—	1.00
AUC	.83	.81	—	.82	—	.57 (NS)
95% CI	.77–.89	.77–.86	—	.77–.87	—	.50–.64
Ages 11–17 (*N* = 324)						
P-YMRS	1.00	.60	.29	.64	.37	.37
P-GBI	.60	1.00	.39	.68	.42	.37
A-GBI	.29	.39	1.00	.36	.54	.28
CBCL	.64	.68	.36	1.00	.53	.46
YSR	.37	.42	.54	.53	1.00	.47
TRF	.37	.37	.28	.46	.47	1.00
AUC	.80	.84	.67	.78	.71	.70
95% CI	.74–.85	.79–.89	.61–.73	.73–.83	.66–.77	.62–.78

Note: AUC = area under the curve; CI = confidence interval; NS = not significant.
All correlations significant *p* < .00005, two tailed, unless otherwise noted.
* *p* < .005, two tailed. AUC CIs based on bootstrapping, 500 resamples with replacement per data file, and adjusted for variance due to imputation.

tendency of teachers to interpret most disruptive behavior as evidence of an attention problem (Abikoff et al., 1993). The higher rate of ADHD in the younger sample combined with teachers' bias toward labeling behaviors as attention problems may contribute to the significantly worse performance of the TRF in the younger sample.

Unique strengths of the current study include the adherence to the 25 recommendations of the STARD guidelines for reporting diagnostic test results (Bossuyt et al., 2003) and the use of relatively large samples, each containing a sizable number of youths (160 and 132 youths) diagnosed with bipolar disorder using the K-SADS as the criterion standard and with the majority of the bipolar diagnoses further confirmed by subsequent psychiatric examination. Another strength is the simultaneous inclusion of four or six different index tests, affording comparisons both between measures and different sources of information (parent, teacher, and youth). The analyses also relied on multiple methods for evaluating diagnostic efficiency, including global estimates and multilevel likelihood ratios, which

provide a clinically meaningful way to interpret test scores. Finally, generalizability of the results was enhanced both by replicating analyses on an independent sample of younger youths (ages 5–10 years) along with the use of bootstrapping procedures to provide nonparametric estimates of confidence intervals that would be less influenced by statistical outliers.

Limitations

Limitations of the study include missing data, both due to protocol changes (the P-YMRS was not available at the outset of data collection) and to difficulty in obtaining teacher data. This limitation was addressed to some extent by performing missing data analyses to identify the size and extent of any patterns associated with partial response and by the use of multiple-imputation procedures—the state of the art method to produce relatively unbiased estimates. A major advantage of the multiple-imputation approach is that it avoids excluding subjects for whom values were missing on an index test, avoiding a potential source for substantial bias in test evaluation (Bossuyt et al., 2003).

TABLE 4
Change in Odds of Bipolar Diagnosis (Likelihood Ratios) for Index Test Scores

Ages 5–10 LR: 50.3% Prevalence of Bipolar Disorders

Summary	Range					
	Low	Mod. Low	Neutral	Mod. High	High	Very High
P-YMRS						
Score	<7	7–13	14–21	22–29	30–34	35+
LR	0.08	0.48	0.88	2.78	6.94	8.92
P-GBI						
Score	<11	11–20	21–30	31–42	43–50	51+
LR	0.10	0.48	1.34	2.31	4.90	6.29
CBCL						
Score	<58	58–67	68–72	73–77	78–81	82+
LR	0.07	0.47	1.50	4.55	3.15	3.52
TRF						
Score	<49	49–56	57–62	63–70	71–77	78+
LR	0.75	0.80	0.88	1.22	1.74	1.28

Ages 11–17 LR: 40.7% Prevalence of Bipolar Disorders

Measure	Range					
	Low	Mod. Low	Neutral	Mod. High	High	Very High
P-YMRS						
Score	<6	6–11	12–17	18–23	24–27	28+
LR	0.20	0.32	0.99	1.99	4.07	7.41
P-GBI						
Score	<9	9–15	16–24	25–39	40–48	49+
LR	0.06	0.25	1.12	2.22	4.82	9.21
A-GBI						
Score	<10	10–17	18–26	27–37	38–45	46+
LR	0.33	1.00	0.83	1.16	2.02	3.92
CBCL						
Score	<54	54–64	65–69	70–75	76–80	81+
LR	0.04	0.53	1.26	2.14	2.65	4.29
YSR						
Score	<49	49–55	56–62	63–69	70–76	77+
LR	0.31	0.52	1.15	1.58	2.32	3.03
TRF						
Score	<46	46–53	54–60	61–68	69–76	77+
LR	0.25	0.64	0.98	2.03	1.47	3.76

Note: LR = likelihood ratio; Low = bottom 20% of sample; Mod. Low = moderately low, 21st to 40th percentile; Neutral = 41st to 60th percentile; Mod. High = moderately high, 61st to 80th percentile; High = 81st to 90th percentile; Very High = top 10%.

Another limitation is that findings are based on samples heavily enriched for bipolar disorder. Likelihood ratios can change when the same index tests are used in different settings (Kraemer, 1992). It will be important to replicate results using same measures in other settings. Also, the current sample included few Hispanic youths, and it will be important for future research to establish whether these measures per-form similarly in Hispanic and other diverse populations.

Finally, it is crucial to note that none of the measures assessed in this study are sufficient for determining a bipolar diagnosis in isolation. These questionnaires were not originally intended to be diagnostic instruments, they do not systematically assess all the features associated with bipolar disorder, and they cannot evalu-

ate cycling, duration, or course of illness. In short, they cannot substitute for a thorough evaluation by a trained professional familiar with the diagnostic criteria for bipolar disorder (American Psychiatric Association, 2001) as well as how the symptoms manifest in children and adolescents (Carlson, 2002; Findling et al., 2001; Geller et al., 2000, 2002b,c; Weckerly, 2002). Similarly, the K-SADS by itself is not considered sufficient to establish a diagnosis of pediatric bipolar disorder (Carlson et al., 2003; Kaufman et al., 1997), even though the rate of agreement between K-SADS diagnosis and clinical diagnosis was high in the current study. In this study, a child and adolescent psychiatrist reviewed the K-SADS protocols and notes to assign a diagnosis for all cases, and any subjects enrolling in a treatment protocol were independently reinterviewed by a child and adolescent psychiatrist.

Clinical Implications

Current findings lead to several concrete clinical recommendations. First, parent report provides powerful information for the recognition and diagnosis of bipolar disorder in youths aged 5 to 17 years, based on the AUC and logistic regression findings. These results contradict the clinical tendency to emphasize self-report when assessing mood disorders (Loeber et al., 1990), and they highlight the need for obtaining collateral sources of information when differentially diagnosing bipolar disorders (Pini et al., 2001). Second, teacher report on the Achenbach TRF does not appear to facilitate the differential diagnosis of juvenile bipolar disorder enough to justify the inconvenience often involved in gathering the information, based on the significantly smaller AUCs and failure to offer any unique information in logistic regression analyses. This might be due to externalizing symptoms not always occurring at school or a general tendency of teachers to attribute any disruptive behavior to an attention deficit/hyperactive disorder (Abikoff et al., 1993), or it might reflect a more narrow limitation of teacher report using the Achenbach instruments (cf. Thuppal et al., 2002). Third, low scores on a parent report on any of the three instruments yield clinically meaningful decreases in the likelihood of a bipolar disorder being present (Table 4); likelihood ratios around 10 or .10 are considered clinically compelling (Sackett et al., 2000). Because the CBCL, P-GBI, and P-YMRS appear roughly equal in their global diagnostic efficiency, there is no need to administer one of the more specialized questionnaires (P-GBI or P-YMRS) if the parent has already endorsed a low level of externalizing behaviors on the CBCL. Fourth, high scores on measures containing manic and hypomanic content provide larger increases in the likelihood of a bipolar diagnosis than do high scores on the Achenbach measures (Table 4). Put another way, the P-YMRS and P-GBI produce fewer false alarms than the Achenbach scales do because high scores are more specific to youths with a bipolar diagnosis. Fifth, there is no clinically meaningful improvement made by combining the scores on the index tests evaluated here, based on the logistic regression analyses. If multiple index tests happen to be available, then the clinician should pick the test with the highest overall diagnostic efficiency or with the most powerful likelihood ratio (based on the values in Table 4). If there are risk factors or clinical concerns that might indicate the presence of a bipolar disorder, then the P-GBI or P-YMRS appears to be the best candidate from the set studied here to provide information changing the likelihood of a bipolar diagnosis to a meaningful degree. If one of the Achenbach instruments already was gathered, then the likelihood ratio resulting from the P-YMRS or the P-GBI should replace the likelihood based on the Achenbach, not be combined with it (Jaeschke et al., 1994; Kraemer, 1992).

Likelihood ratios can be applied as a change in odds of a diagnosis. Clinicians can use the change in odds by converting a probability into odds and then multiplying the odds by the likelihood ratio. For example, if a clinician estimated that there was a 10% chance of a child having a bipolar diagnosis before looking at the test result (deriving either from clinical impressions or from the base rate of bipolar disorder at that particular setting), then the initial odds of a bipolar diagnosis would be 0.10/0.90 = 0.11. If the parent reported a score of 26 on the P-YMRS, then the change in odds for a child younger than age 11 would be 2.78, based on Table 4. The new odds would be 0.11 times 2.78, or 0.31, which translates back into a 23.4% probability. Clinicians can combine a previous probability with a test result using a visual nomogram, analogous to a slide rule that allows one to connect two dots to read off the new probability without having to perform any calculations, and Excel spreadsheets and personal digital assistant programs are available to make it practical to use likelihood ratios clinically (Sackett et al., 2000).

The likelihood ratios presented in Table 4 are likely to be most useful in two situations. One is helping to clarify difficult or ambiguous diagnostic presentations. The P-YMRS or P-GBI are capable of moving the probability of a bipolar diagnosis from a pretest probability of 50% to either less than 10% or approaching 90% in the event of extremely low or high scores (see Jaeschke et al., 1994; Sackett et al., 2000 for detailed examples). In situations in which the clinician is uncertain about the differential diagnosis, these tests can contribute by virtue of systematically assessing relevant symptoms in a standardized fashion, and then by providing empirical estimates of changes in likelihood of a bipolar diagnosis (Meehl, 1954).

The second potential benefit of applying the likelihood ratios is that it could substantially decrease the rate of false-positive diagnoses in settings in which bipolar disorder is likely to be rare. Low base rate situations, such as community mental health and general practice settings, make it difficult for clinicians to interpret risk factors optimally (Dawes et al., 1989). The actuarial use of one of the parent-report index tests (P-GBI, P-YMRS, or CBCL) would often help to correctly assign a low risk of bipolar disorder to a case where clinicians might be prone to overestimate risk due to cognitive biases and heuristics (Davidow and Levinson, 1993).

Again, we stress that these posterior probabilities are not the same thing as a diagnosis. The diagnosis of bipolar disorder is a high-stakes decision that requires the careful clinical evaluation of mood symptoms, including attention to intensity and duration, as well as a thorough search for evidence of mood cycling (Findling et al., 2001; Geller et al., 1995). Current findings indicate that several inexpensive and convenient parent-completed rating scales could facilitate accurate diagnosis, particularly by reducing the number of false-positive diagnoses in children and adolescents seen at outpatient and community settings. These tests can contribute to the assessment process by raising "red flags" when high scores occur during an initial assessment or screening, indicating when more specialized evaluation is warranted. Low scores on parent measures are also more decisive in helping "rule out" bipolar disorder, even in fairly ambiguous situations.

REFERENCES

Abikoff H, Courtney M, Pelham WE, Koplewicz HS (1993), Teachers' ratings of disruptive behaviors: the influence of halo effects. *J Abnorm Child Psychol* 21:519–533

Achenbach TM (1991a), *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont

Achenbach TM (1991b), *Manual for the Teacher's Report Form and 1991 Profile* Burlington: University of Vermont

Achenbach TM (1991c), *Manual for the Youth Self-Report Form and 1991 Profile* Burlington: University of Vermont

American Psychiatric Association (1994), *Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV)*.Washington, DC: American Psychiatric Association

American Psychiatric Association (2001), *Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association

Berument SK, Rutter M, Lord C, Pickles A, Bailey A (1999), Autism screening questionnaire: diagnostic validity. *Br J Psychiatry* 175:444–451

Biederman J, Klein RG, Pine DS, Klein DF (1998), Resolved: mania is mistaken for ADHD in prepubertal children. *J Am Acad Child Adolesc Psychiatry* 37:1091–1099

Biederman J, Mick E, Faraone SV, Spencer T, Wilens TE, Wozniak J (2000), Pediatric mania: a developmental subtype of bipolar disorder? *Biol Psychiatry* 48:458–466

Biederman J, Wozniak J, Kiely K et al. (1995), CBCL clinical scales discriminate prepubertal children with structured interview-derived diagnosis of mania from those with ADHD. *J Am Acad Child Adolesc Psychiatry* 34:464–471

Bird HR, Gould MS, Staghezza B (1992), Aggregating data from multiple informants in child psychiatry epidemiological research. *J Am Acad Child Adolesc Psychiatry* 31:78–85

Bossuyt PM, Reitsma JB, Bruns DE et al. (2003), Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med* 138:40–44

Bowring MA, Kovacs M (1992), Difficulties in diagnosing manic disorders among children and adolescents. *J Am Acad Child Adolesc Psychiatry* 31:611–614

Carlson GA (1998), Mania and ADHD: Comorbidity or confusion. *J Affect Disord* 51:177–187

Carlson GA (2002), Bipolar disorder in children and adolescents: a critical review. In: *The Many Faces of Depression in Children and Adolescents, Vol. 21,* Shaffer D, Waslick B, eds. Washington, DC: APPI Press, pp 105–128

Carlson GA, Jensen PS, Findling RL et al. (2003), Methodological issues and controversies in clinical trials with child and adolescent patients with bipolar disorder: report of a consensus conference. *J Child Adolesc Psychopharmacol* 13:1–15

Carlson GA, Kelly KL (1998), Manic symptoms in psychiatrically hospitalized children: what do they mean? *J Affect Disord* 51:123–135

Carlson GA, Loney J, Salisbury H, Volpe RJ (1998), Young referred boys with DICA-P manic symptoms vs. two comparison groups. *J Affect Disord* 121:113–121

Danielson CK, Youngstrom EA, Findling RL, Calabrese JR (2003), Discriminative validity of the General Behavior Inventory using youth report. *J Abnorm Child Psychol* 31:29–39

Davidow J, Levinson EM (1993), Heuristic principles and cognitive bias in decision making: implications for assessment in school psychology. *Psychol Sch* 30:351–361

Dawes RM, Faust D, Meehl PE (1989), Clinical versus actuarial judgment. *Science* 243:1668–1674

Depue RA, Krauss S, Spoont MR, Arbisi P (1989), General Behavior Inventory identification of unipolar and bipolar affective conditions in a nonclinical university population. *J Abnorm Psychol* 98:117–126

Depue RA, Slater JF, Wolfstetter-Kausch H, Klein DN, Goplerud E, Farr DA (1981), A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: a conceptual framework and five validation studies. *J Abnorm Psychol* 90:381–437

Dienes KA, Chang KD, Blasey CM, Adleman N, Steiner H (2002), Characterization of children of bipolar parents by parent report CBCL. *J Psychiatr Res* 36:337–345

Drotar D, Stein REK, Perrin EC (1995), Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. In: *Methodological Issues in Clinical Child Psychology Research. J Clin Child Psychol* 24(special issue)184–192

Dunn LM, Dunn LM (1997), *Examiner's Manual for the Peabody Picture Vocabulary Test—Third Edition*. Circle Pines, MN: American Guidance Service

Efron B (1982), *The Jackknife, the Bootstrap, and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics

Findling RL, Gracious BL, McNamara NK, Youngstrom EA, Demeter C, Calabrese JR (2001), Rapid, continuous cycling and psychiatric co-morbidity in pediatric bipolar I disorder. *Bipolar Disord* 3:202–210

Findling RL, Youngstrom EA, Danielson CK et al. (2002), Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. *Bipolar Disord* 4:34–42

Frazier JA, Biederman J, Tohen M et al. (2001), A prospective open-label treatment trial of olanzapine monotherapy in children and adolescents with bipolar disorder. *J Child Adolesc Psychopharmacol* 11:239–250

Geller B, Craney JL, Bolhofner K, Nickelsburg MJ, Williams M, Zimerman B (2002a), Two-year prospective follow-up of children with a prepubertal and early adolescent bipolar disorder phenotype. *Am J Psychiatry* 159:927–933

Geller B, Luby J (1997), Child and adolescent bipolar disorder: a review of the past 10 years. *J Am Acad Child Adolesc Psychiatry* 36:1168–1176

Geller B, Sun K, Zimerman B, Luby J, Frazier J, Williams M (1995), Complex and rapid-cycling in bipolar children and adolescents: a preliminary study. *J Affect Disord* 34:259–268

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

Geller B, Zimerman B, Williams M et al. (2000), Diagnostic characteristics of 93 cases of prepubertal and early adolescent bipolar disorder phenotype by gender, puberty and comorbid attention deficit hyperactivity disorder. *J Child Adolesc Psychopharmacol* 10:157–164

Geller B, Zimerman B, Williams M et al. (2002b), *DSM-IV* mania symptoms in a prepubertal and early adolescent bipolar disorder phenotype compared to attention-deficit hyperactive and normal controls. *J Child Adolesc Psychopharmacol* 12:11–25

Geller B, Zimerman B, Williams M, Delbello MP, Frazier J, Beringer L (2002c), Phenomenology of prepubertal and early adolescent bipolar disorder: examples of elated mood, grandiose behaviors, decreased need for sleep, racing thoughts and hypersexuality. *J Child Adolesc Psychopharmacol* 12:3–9

Gracious BL, Youngstrom EA, Findling RL, Calabrese JR (2002), Discriminative validity of a parent version of the Young Mania Rating Scale. *J Am Acad Child Adolesc Psychiatry* 41:1350–1359

Graham JW, Schafer JL (1999), On the performance of multiple imputation for multivariate data with small sample size. In: *Statistical Strategies for Small Sample Research*, Hoyle RH, ed. Thousand Oaks, CA: Sage, pp 1–29

Hanley JA, McNeil BJ (1983), A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843

Hazell PL, Lewin TJ, Carr VJ (1999), Confirmation that Child Behavior Checklist clinical scales discriminate juvenile mania from attention deficit hyperactivity disorder. *J Paediatr Child Health* 35:199–203

Hellander M (2002), *Lithium testing in children: a public health necessity.* Washington, DC: Testimony to the U.S. Food and Drug Administration (available at www.cabf.org)

Hirschfeld RMA, Williams JBW, Spitzer RL et al. (2000), Development and validation of a screening instrument for bipolar spectrum disorder: the mood disorder questionnaire. *Am J Psychiatry* 157:1873–1875

Hodgins S, Faucher B, Zarac A, Ellenbogen M (2002), Children of parents with bipolar disorder. A population at high risk for major affective disorders. *Child Adolesc Psychiatr Clin N Am* 11:533–553

Hosmer DW, Lemeshow S (1989), *Applied Logistic Regression.* New York: Wiley

Jaeschke R, Guyatt GH, Sackett DL (1994), 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* 271:389–391

Kafantaris V (1995), Treatment of bipolar disorder in children and adolescents. *J Am Acad Child Adolesc Psychiatry* 34:732–741

Kahana SY, Youngstrom EA, Findling RL, Calabrese JR (2003), 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* 13:471–488

Kaufman J, Birmaher B, Brent D et al. (1997), 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* 36:980–988

Kim EY, Miklowitz DJ (2002), Childhood mania, attention deficit hyperactivity disorder and conduct disorder: a critical review of diagnostic dilemmas. *Bipolar Disord* 4:215–225

Kowatch RA, Suppes T, Carmody TJ et al. (2000), Effect size of lithium, divalproex sodium, and carbamazepine in children and adolescents with bipolar disorder. *J Am Acad Child Adolesc Psychiatry* 39:713–720

Kraemer HC (1992), *Evaluating Medical Tests: Objective and Quantitative Guidelines.* Newbury Park, CA: Sage

Leibenluft E, Charney DS, Towbin KE, Bhangoo RK, Pine DS (2003), Defining clinical phenotypes of juvenile mania. *Am J Psychiatry* 160:430–437

Lewinsohn PM, Seeley JR, Klein DN (2003), Bipolar disorder in adolescents: Epidemiology and suicidal behavior. In: *Bipolar Disorder in Childhood and Early Adolescence,* Geller B, DelBello MP, eds. New York: Guilford, pp 7–24

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

Loeber R, Green SM, Lahey BB, Stouthamer Loeber M (1989), Optimal informants on childhood disruptive behaviors. *Dev Psychopathol* 1:317–337

Meehl PE (1954), *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* Minneapolis: University of Minnesota Press

Naylor MW, Anderson TR, Kruesi MJ, Stoewe M (2002), Pharmacoepidemiology of bipolar disorder in abused and neglected state wards. Presented at the Annual Meeting of the American Academy of Child and Adolescent Psychiatry, San Francisco, October

Nottelmann E, Biederman J, Birmaher B et al. (2001), National Institute of Mental Health research roundtable on prepubertal bipolar disorder. *J Am Acad Child Adolesc Psychiatry* 40:871–878

Pini S, Cassano GB, Dell'Osso L, Amador XF (2001), Insight into illness in schizophrenia, schizoaffective disorder, and mood disorders with psychotic features. *Am J Psychiatry* 158:122–125

Pliszka SR, Sherman JO, Barrow MV, Irick S (2000), Affective disorder in juvenile offenders: a preliminary study. *Am J Psychiatry* 157:130–132

Pogge DL, Wayland-Smith D, Zaccario M, Borgaro S, Stokes J, Harvey PD (2001), Diagnosis of manic episodes in adolescent inpatients: structured diagnostic procedures compared to clinical chart diagnoses. *Psychiatry Res* 101:47–54

Poznanski EO, Miller E, Salguero C, Kelsh RC (1984), Preliminary studies of the reliability and validity of the Children's Depression Rating Scale. *J Am Acad Child Psychiatry* 23:191–197

Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB (2000), *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd ed. New York: Churchill Livingstone

Sattler JM (2002), *Assessment of Children: Behavioral and Clinical Applications*, 4th ed. San Diego: J.M. Sattler

Schafer JL (2002), Multiple imputation with PAN. In: *New Methods for the Analysis of Change*, Sayer AG, ed. Washington, DC: American Psychological Association, pp 355–378

Thuppal M, Carlson GA, Sprafkin J, Gadow KD (2002), Correspondence between adolescent report, parent report, and teacher report of manic symptoms. *J Child Adolesc Psychopharmacol* 12:27–35

Wagner KD, Weller EB, Carlson GA et al. (2002), An open-label trial of divalproex in children and adolescents with bipolar disorder. *J Am Acad Child Adolesc Psychiatry* 41:1224–1230

Weckerly J (2002), Pediatric bipolar mood disorder. *J Dev Behav Pediatr* 23:42–56

Wozniak J, Biederman J, Monteaux M, Richards J, Faraone SV (2002), Parsing the comorbidity between bipolar disorder and anxiety disorders: a familial risk analysis. *J Child Adolesc Psychopharmacol* 12:101–111

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

Youngstrom EA, Findling RL, Danielson CK, Calabrese JR (2001), Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. *Psychol Assess* 13:267–276

Youngstrom EA, Gracious BL, Danielson CK, Findling RL, Calabrese JR (2003), Toward an integration of parent and clinician report on the Young Mania Rating Scale. *J Affect Disord* 77:179–190