

Discriminative Validity of a Parent Version of the Young Mania Rating Scale

BARBARA L. GRACIOUS, M.D., ERIC A. YOUNGSTROM, PH.D., ROBERT L. FINDLING, M.D.,
AND JOSEPH R. CALABRESE, M.D.

ABSTRACT

Objective: To examine the usefulness of a parent report version of the Young Mania Rating Scale (P-YMRS) in distinguishing bipolar disorder from other mental health conditions in children and adolescents. **Method:** Parents of 117 youths aged 5 to 17 years presenting to an outpatient research center completed an adapted Young Mania Rating Scale (P-YMRS). Eligible subjects underwent a diagnostic evaluation including a semistructured instrument (Schedule for Affective Disorders and Schizophrenia for School-Age Children) and also a clinical evaluation by a child and adolescent psychiatrist in more than 75% of the subjects. **Results:** Factor analyses of the P-YMRS suggested one dimension, with a total score showing acceptable internal consistency ($\alpha = .75$). Logistic regressions discriminated bipolar mood disorder versus unipolar disorder, versus disruptive behavior disorder, and versus any other diagnosis. Classification rates exceeded 78%, and receiver operating characteristics analyses showed good diagnostic efficiency, with areas under the curve greater than 0.82. **Conclusions:** The P-YMRS may be used to derive clinically meaningful information about mood disorders in youths. *J. Am. Acad. Child Adolesc. Psychiatry*, 2002, 41(11):1350–1359. **Key Words:** bipolar disorder, Young Mania Rating Scale.

It is often clinically difficult to differentiate bipolar disorder from other mental health conditions in children and adolescents, especially attention-deficit/hyperactivity disorder (ADHD), conduct disorder, and schizophrenia (Bowring and Kovacs, 1992; Fristad et al., 1992a,b; Weller et al., 1986). An easy-to-use, accurate rating scale would aid clinicians and researchers in this process, potentially reducing the time to a correct diagnosis, reducing overdiagnosis, increasing the likelihood of appropriate treatment with mood stabilizers, and enhancing the feasibility of epidemiological studies of the incidence and prevalence of bipolar disorder in children and adolescents.

The Young Mania Rating Scale (Y-MRS) was developed by Young and colleagues (1978) to be used by a

trained clinician during a 15- to 30-minute patient interview. A severity score ranging from 0 (symptoms not present/normal behavior) to 4 to 8 (extreme deviance) for each question is based on “the patient’s subjective report of his/her condition over the past 48 hours and the clinician’s behavioral observations during the interview, with emphasis on the latter” (p. 433). Acceptable validity and reliability was demonstrated in 20 adults rated during their first week in the hospital, with a correlation of 0.93 between raters (Young et al., 1978). Young et al. stated that “the scale is not intended to be used as a diagnostic instrument. It appears to measure the manic ‘state’ as opposed to traits, since there was virtually no correlation between scores in individuals rated before and after two weeks of treatment” (p. 433). The earliest publication about outpatient use of the Y-MRS (in adults) found via Medline search is Hatterer et al. (1988; see Poolsup et al., 1999, for review). Little has been published on how the Y-MRS should be used in the outpatient setting, however, which results in differences between research groups in how raters are interpreting both items and symptom severity. In addition, information about its use in outpatients appears to be largely communicated through unpublished means. The following paragraph summarizes the published data on outpatient use in children and adolescents to date.

Accepted May 28, 2002.

From Case Western Reserve University, Cleveland (B.L.G., E.A.Y., R.L.F., J.R.C.), and the University Hospitals of Cleveland (B.L.G., R.L.F., J.R.C.).

This research was supported by a Clinical Research Center Grant from the Stanley Foundation. The authors thank Lisa Branicky, Raisa David, Christine Demeter, Denise DelPorto, Lisa Townsend, Cara West, and Resa Whipkey for their assistance with this project, as well as all the families that participated.

Correspondence to Dr. Gracious, Department of Psychiatry, Srong Memorial Hospital, University of Rochester Medical Center, 300 Crittenden Blvd., Rochester, NY 14642; e-mail: Barbara_Gracious@URMC.Rochester.edu.

0890-8567/02/4111-1350©2002 by the American Academy of Child and Adolescent Psychiatry.

DOI: 10.1097/01.CHL.0000024841.60748.9A

Fristad and colleagues (1992b) examined the ability of the Y-MRS to distinguish prepubertal children with mania versus ADHD in 11 inpatient subjects. Validity of the Y-MRS in children was demonstrated by comparing the Y-MRS with *DSM-III-R* criteria, as well as with ADHD rating scales and a structured interview. Fristad et al. (1995) studied 30 additional prepubertal subjects and found the Y-MRS to have adequate internal consistency ($\alpha = .80$), convergent validity ($r = 0.83, p < .0001$), and divergent validity (no significant correlations with depression and hyperactivity ratings). Items assessing "classic" manic symptoms such as elevated mood, increased sexual interest, pressured speech, and racing thoughts discriminated the bipolar group from two comparison groups. Items assessing increased activity level and irritability did not. Gracious et al. (1994) obtained an interrater reliability of 0.86 from 10 taped subject interviews taken from a group of 108 outpatient children and adolescents aged 3 to 18 years. Excellent test-retest reliability for a single rater was dependent on standardizing definitions of item terms and interview questions for nonvisual items such as insight and sleep. Gracious et al. (1994) then adapted and pilot-studied a parent version of the Y-MRS (P-YMRS), finding an intraclass correlation coefficient of 0.86 for the P-YMRS in the group of 108 subjects. Intraclass correlation coefficients for 13 mother-father pairs were 0.49 and 0.50 for before and after the appointment, consistent with results from mother-father pairs for other rating scales.

Specific research questions concerning the P-YMRS include the following: (1) When parents complete the P-YMRS to describe their child's manic symptoms, does the result conform to a simple, one-factor solution? (2) Does the instrument meet established psychometric criteria for reliability? (3) Do youths with different *DSM-IV* diagnoses show significant mean differences on parent-reported levels of manic symptoms? Youths with bipolar spectrum disorders should score higher on a measure of manic symptoms than youths with just unipolar mood, disruptive behavior, or other disorders. If the P-YMRS demonstrates substantial separations between group means, this would indicate that construction of thresholds for defining clinically significant change may be appropriate, such as when an individual's score could move out of a clinical distribution and into a nonclinical range (e.g., Jacobson and Truax, 1991). (4) How well do parent ratings on the P-YMRS converge with diagnoses of bipolar mood disorder based on structured clin-

ical interviews with the child, since the P-YMRS and the diagnosis rely on different sources of information (parent questionnaire versus clinical interview of the parent and youth)? A rating form considered for inclusion in an assessment battery needs to demonstrate this kind of validity to have practical value (Wiggins, 1973; Youngstrom and Drotar, unpublished, 2002).

METHOD

Participants

Eligible patients were identified and referred from two distinct pediatric psychopharmacology research infrastructures at a single Midwestern urban outpatient research clinic specializing in the treatment of mood disorders. The Institutional Review Board for Human Investigation of the University Hospitals of Cleveland approved the protocol. Participants included 117 children and adolescents and their parents or guardians. Recruitment occurred from August 17, 1999, to November 3, 2000. Sixty-six (56.4%) were male. Eighty-four percent were white, 3.6% were African American, 5.5% were Hispanic, and 3.6% were of other ethnicity. Subjects ranged in age from 5 to 17 years (mean = 12.4, SD = 3.1). Children and adolescents with a psychiatric disorder due to a general medical condition, pervasive developmental disorder, or evidence of mental retardation were excluded. All parents read English adequately to complete the instrument.

Measures

The P-YMRS rating form has 11 multiple-choice items that are scored from 0 to 8 as described above for the Y-MRS, with a total score calculated. A comparison of the anchor items for each question of the Y-MRS versus the P-YMRS is presented in Table 1. Completion time for parents is approximately 5 minutes. The semistructured Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS) (Kaufman et al., 1997) was administered as described below to assess diagnoses.

Procedure

After informed consent for a screening assessment, the primary caregiver completed the P-YMRS at the same visit time as the research diagnostic interview of the child, conducted by the research assistant. The custodial mother (90%) or father (10%) completed the P-YMRS. Primary diagnoses of the children and adolescents were obtained from either the K-SADS-Epidemiologic version (K-SADS-E) or the Present and Lifetime version (K-SADS-PL) (Kaufman et al., 1997), completed by research assistants with either bachelor's ($n = 4$) or master's level degrees ($n = 3$). Both the child and the parent were administered the K-SADS separately by the same interviewer during individual interviews. If the youth met inclusion/exclusion criteria for a clinical trial, diagnoses were confirmed by an independent clinical evaluation by a child and adolescent psychiatrist. In situations of diagnostic uncertainty after a K-SADS interview, a consensus diagnosis was made after the assessment by the child and adolescent psychiatrist. Both a research assistant and a child and adolescent psychiatrist interviewed 78% of subjects, with rare disagreement on diagnoses. Research assistants were trained to criterion (overall κ of >0.85 on each evaluation) by conducting five K-SADS interviews with an experienced rater observing the interview and scoring independently. Research assistants observed an average of 10 to 15 K-SADS interviews before they

TABLE 1
Anchor Item Comparisons Between the Y-MRS and P-YMRS

Y-MRS Item Descriptor	P-YMRS Item Descriptor	Y-MRS Scoring Descriptors	P-YMRS Scoring Descriptors
1. Elevated Mood	Is your child's mood higher (better) than usual?	0. Absent 1. Mildly or possibly increased on questioning 2. Definite subjective elevation; optimistic, self confident; cheerful; appropriate to content 3. Elevated; inappropriate to content; humorous 4. Euphoric; inappropriate laughter; singing	0. No 1. Mildly or possibly increased 2. Definite elevation—more optimistic, self confident; cheerful; appropriate to their conversation 3. Elevated but inappropriate to content; joking, mildly silly 4. Euphoric; inappropriate laughter; singing/making noises; very silly
2. Increased Motor Activity/Energy	Does your child's energy level or motor activity appear to be greater than usual?	0. Absent 1. Subjectively increased 2. Animated; gestures increased 3. Excessive energy; hyperactive at times (can be calmed) 4. Motor excitement; continuous hyperactivity (cannot be calmed)	0. No 1. Mildly or possibly increased 2. More animated; increased gesturing 3. Energy is excessive 4. Very excited; continuous hyperactivity; cannot be calmed
3. Sexual Interest	Is your child showing more than usual interest in sexual matters?	0. Normal, not increased 1. Mildly or possibly increased 2. Definite subjective increase on question 3. Spontaneous sexual content; elaborates on sexual matters; hypersexual by self-report 4. Overt sexual acts (toward patients, staff, or interviewer)	0. No 1. Mildly or possibly increased 2. Definite increase when the topic arises 3. Talks spontaneously about sexual matters; gives more detail than usual 4. Has shown open sexual behavior—touching others or self inappropriately
4. Sleep	Has your child's sleep decreased lately?	0. Reports no decrease in sleep 1. Sleeping less than normal amount by up to 1 hour 2. Sleeping less than normal by more than 1 hour 3. Reports decreased need for sleep 4. Denies need for sleep	0. No 1. Sleeping less than normal amount by up to 1 hour 2. Sleeping less than normal amount by more than 1 hour 3. Need for sleep appears decreased; less than 4 hours 4. Denies need for sleep; has stayed up one night or more
5. Irritability	Has your child appeared irritable?	0. Absent 2. Subjectively increased 4. Irritable at times during interview; recent episodes of anger or annoyance 6. Frequently irritable during interview; short, curt throughout 8. Hostile, uncooperative; interview impossible	0. No more than usual 2. More grouchy or crabby 4. Irritable openly several times throughout the day; recent episodes of anger with family, at school, or with friends 6. Frequently irritable to point of being rude or withdrawn 8. Hostile and uncooperative about all the time
6. Speech (Rate and Amount)	Is your child talking more quickly or more than usual?	0. No increase 2. Feels talkative 4. Increased rate or amount at times, verbose at times 6. Push; consistently increased rate and amount; difficult to interrupt 8. Pressured; uninteruptible, continuous speech	0. No change 2. Seems more talkative 4. Talking faster or more to say at times 6. Talking more or faster to point he/she is difficult to interrupt 8. Continuous speech; unable to interrupt

—Continued

TABLE 1 (Continued)
Anchor Item Comparisons Between the Y-MRS and P-YMRS

Y-MRS Item Descriptor	P-YMRS Item Descriptor	Y-MRS Scoring Descriptors	P-YMRS Scoring Descriptors
7. Language-Thought Disorder	Has your child shown changes in his/her thought patterns?	0. Absent 1. Circumstantial; mild distractibility; quick thoughts 2. Distractible; loses goal of thought; changes topics frequently 3. Flight of ideas; tangentiality; difficult to follow 4. Incoherent; communication is impossible	0. No 1. Thinking faster; some decrease in concentration; talking "around the issue" 2. Distractible; loses track of the point; changes topics frequently; thoughts racing 3. Difficult to follow; goes from one idea to the next; topics do not relate; makes rhymes or repeats words 4. Not understandable; he/she doesn't seem to make any sense
8. Content	Is your child talking about different things than usual?	0. Normal 2. Questionable plans; new interests 4. Special project(s); hyperreligious 6. Grandiose or paranoid ideas; ideas of reference 8. Delusions; hallucinations	0. No 2. He/she has new interests and is making more plans 4. Making special projects; more religious or interested in God 6. Thinks more of him/herself; believes he/she has special powers; believes he/she is receiving special messages 8. Is hearing unreal noises/voices; detects odors no one else smells; feels unusual sensations; has unreal beliefs
9. Disruptive/Aggressive	Has your child been more disruptive or aggressive?	0. Absent, cooperative 2. Sarcastic; loud at times, guarded 4. Demanding, threats on ward 6. Threatens interviewer, shouting, interview difficult 8. Assaultive; destructive, interview impossible	0. No; he/she is cooperative 2. Sarcastic; loud; defensive 4. More demanding; making threats 6. Has threatened a family member or teacher; shouting; knocking over possessions/furniture or hitting a wall 8. Has attacked family member, teacher, or peer; destroyed property; cannot be spoken to without violence
10. Appearance	Has your child's interest in his/her appearance changed recently?	0. Appropriate dress and grooming 1. Minimally unkempt 2. Poorly groomed; moderately disheveled, overdressed 3. Disheveled; partially clothed; garish makeup 4. Completely unkempt; decorated; bizarre garb	0. No 1. A little less or more interest in grooming than usual 2. Doesn't care about washing or changing clothes, or is changing clothes more than three times a day 3. Very messy; needs to be supervised to finish dressing; applying makeup in overly-done or poor fashion 4. Refuses to dress appropriately; wearing bizarre styles
11. Insight	Does your child think he/she needs help at this time?	0. Present; admits illness; agrees with need for treatment 1. Possibly ill 2. Admits behavior change, but denies illness 3. Admits possible change behavior, but denies illness 4. Denies any behavior change	0. Yes; admits difficulties and wants treatment 1. Believes there might be something wrong 2. Admits to change in behavior but denies he/she needs help 3. Admits behavior might have changed but denies need for help 4. Denies there have been any changes in his/her behavior/thinking

Note: Y-MRS = Young Mania Rating Scale; P-YMRS = parent version of the Y-MRS.

began independent scoring for interrater reliability. Acceptable interrater reliability ($\kappa > 0.85$) was maintained by joint rating sessions at every 10th interview. The researcher performing the K-SADS interview did not see the P-YMRS during the diagnostic process.

Statistics

We tested the overall efficiency of the P-YMRS as well as the contributions of each scale by using logistic regression to predict diagnostic category. We also evaluated the receiver operating characteristics (ROC) of the P-YMRS for making different diagnostic decisions, because this technique provides a metric for directly comparing the information value of competing measures in terms of their diagnostic sensitivity and specificity (McFall and Treat, 1999). Both sensitivity and specificity are theoretically independent of base rates and are therefore more likely to generalize to other samples than other diagnostic efficiency statistics, such as positive or negative predictive power (Baldessarini et al., 1983; Ingelfinger et al., 1994; Kraemer, 1992).

RESULTS

Diagnoses

Primary diagnoses and rates of comorbidity for ADHD are presented in Table 2. Ten (55.5%) of the 18 youths categorized with a disruptive disorder or ADHD also had a comorbid unipolar mood disorder. The central issue concerning the P-YMRS is whether it detects bipolar spec-

trum disorders, regardless of comorbid conditions. Subjects with bipolar spectrum disorders were categorized together whether or not they also presented with comorbid attention problems. Subjects meeting criteria for both ADHD and a unipolar mood disorder were assigned to the “disruptive behavior” category, reflecting the fact that ADHD is more difficult than a pure unipolar presentation to differentiate from manic and hypomanic states.

Preliminary Analyses

Data Completion. Complete data on all P-YMRS items were available for 105 cases. The remaining 12 cases were missing one item. Only cases with complete data on all items were included in the reliability and factor analyses. Scale scores were computed for all cases, even if they omitted one item. The most frequently omitted item was number 11, the “insight” into the disorder item, which was skipped seven times. No other item was omitted more than twice. There were no demographic or diagnostic differences between those with complete versus partial P-YMRS data (all bivariate p values $> .05$).

Factor Analyses. Exploratory factor analyses tested the adequacy of the one-dimension, total score format used with the MRS. Two techniques—Horn’s parallel analysis and the minimum average partial (MAP) method—have demonstrated substantially better accuracy than other decision rules (Velicer et al., 2000). Both were applied to the present data, and both indicated that a one-factor solution best describes the pattern of correlations among the items.

Table 3 presents the item means, standard deviations, factor loadings, and corrected item-total correlations. Factor loadings examine the correlation between the score on a particular item and the subject’s latent factor of mania. The corrected item-total correlation is the correlation between a particular item and the total of all other items. Based on both factor loadings and the corrected item-total correlations, three items do not contribute significantly to the total score in our sample: sexual interest (item 3), content (i.e., grandiosity, delusions; item 8), and insight (item 11). Using conventional guidelines for scale development, these items could be excluded (Streiner and Norman, 1995). The resulting 8-item scale correlates highly with the 11-item version ($r = 0.97$).

Reliability and Reliable Change Indices. A total score on the 8-item P-YMRS demonstrated moderate internal consistency (Cronbach $\alpha = .75$; $.72$ for the 11-item version), with a standard error of measurement of 3.42 points (4.29

TABLE 2

Primary Diagnoses and Rates of Comorbidity for ADHD

Group	n	%	Comorbid ADHD	
			n	%
Bipolar I	39	33.3	30	76.9
Unipolar mood	31	26.8	NA ^a	NA ^a
Major depressive disorder	25	21.4	NA ^a	NA ^a
Dysthymia	4	3.4	NA ^a	NA ^a
Depressive Disorder NOS	2	1.7	NA ^a	NA ^a
Other bipolar	19	16.2	10	52.6
Bipolar II	2	1.7	0	0
Cyclothymia	4	3.4	2	50.0
Bipolar NOS	13	11.1	8	61.5
Disruptive behavior disorders	18	15.4	1	5.6
ADHD-combined type	8	6.8	NA	NA
ADHD-inattentive type	4	3.4	NA	NA
ADHD-NOS	3	2.6	NA	NA
Oppositional defiant disorder	1	0.9	0	0
Conduct disorder	1	0.9	0	0
Disruptive behavior disorder NOS	1	0.9	1	100
Residual	10	8.5	0	0
Post-traumatic stress disorder	4	3.4	0	0
Schizophrenia/schizoaffective	3	2.6	0	0
No diagnosis	3	2.6	0	0

Note: ADHD = attention-deficit/hyperactivity disorder; NOS = not otherwise specified; NA = not applicable.

^a Subjects meeting criteria for both ADHD and a unipolar mood disorder were categorized as having ADHD.

TABLE 3

Descriptive Statistics, Factor Loadings, and Corrected Item-Total Correlations for P-YMRS items ($n = 105$)

P-YMRS Item	Mean	SD	Factor Loading	Corrected Item-Total r
1. Elevated Mood	1.35	1.52	0.51	0.38
2. Increased Motor Activity/Energy	1.46	1.57	0.73	0.58
3. Sexual Interest	0.53	1.05	0.28	0.23
4. Sleep	0.83	1.21	0.48	0.38
5. Irritability	3.86	2.46	0.45	0.43
6. Speech (Rate and Amount)	2.16	2.49	0.58	0.44
7. Thoughts	1.03	1.17	0.61	0.50
8. Content	0.87	1.99	0.22	0.16
9. Disruptive/Aggressive Behavior	3.44	2.75	0.56	0.55
10. Appearance	0.76	1.03	0.33	0.30
11. Insight	1.53	1.53	0.06	0.06

Note: Factor loadings based on principal axis factoring extracting one factor. P-YMRS = parent version of the Young Mania Rating Scale.

points for the 11-item). Confidence intervals and critical scores for reliable change use the standard error of the difference. Individual changes of 10 points or more on the 8-item P-YMRS (12 points on the 11-item P-YMRS) are 95% likely to reflect a statistically reliable change in manic symptoms (i.e., z of $1.96 \times$ standard error of difference of $4.83 = 9.47$, rounded up to 10). To calculate the reliable change index (RCI) proposed by Jacobson and Truax (1991), one would take the difference between the case's two P-YMRS scores and divide it by the value for the standard error of the difference. For example, a person with an intake score of 33 on the 8-item P-YMRS and a score of 22 at follow-up would have an RCI of 2.28 ($RCI = [33 - 22]/4.83$). RCIs greater than 1.65 are considered 90% reliable. RCIs greater than 1.96 are 95% likely to reflect real change versus measurement error.

Discriminative Validity. We used two different approaches to examine the discriminative validity of the P-YMRS scales. The first, more commonly used method involved categorizing the group based on primary Axis I diagnosis, as derived from a structured clinical interview, and then using analysis of variance to determine which groups differed on average P-YMRS scores. The second approach used logistic regression to determine the relative value of the two scales to make different diagnostic distinctions and evaluate the overall classification accuracy based on the P-YMRS. This approach is less widely used, but it addresses the question of greater clinical importance: How helpful is the P-YMRS in making differential diagnoses?

Group Differences on P-YMRS. Table 4 presents the group means for the five diagnostic clusters. The five groups showed significant differences on the P-YMRS total score, regardless of whether it was based on all 11 items ($F_{4,112} = 17.69, p < .0005$) or 8 items ($F_{4,112} = 19.94, p < .0005$). Using the Tukey honestly significant difference test to determine reliable post hoc group differences, the bipolar spectrum disorders scored significantly higher than did the unipolar depression and residual diagnostic groups. Bipolar I participants scored significantly higher than did youths with ADHD, but ADHD and other bipolar spectrum disorders (bipolar II, cyclothymia, and bipolar-not otherwise specified) were not reliably different (all p values $< .0005$). Ten of 18 participants with ADHD also had a comorbid diagnosis of unipolar depression or dysthymia. These youths can be difficult to discern from bipolar spectrum individuals, because both can display impulsivity and heightened energy as well as periods of depressed mood. However, the comorbid youths did not score significantly different from the youths with ADHD and no mood disorder (11-item scale: mean = 14.3 versus 12.4, not significant [NS]; 8-item scale: 12.4 versus 10.5, NS).

Because of the large range of ages included in the sample, scores on the P-YMRS total and individual items were compared for children (aged < 13) and teenagers (aged $13+$). The younger age group scored significantly higher on the P-YMRS total for both the standard 11-item version and the 8-item version (t_{115} values = 4.10 and 3.12, respectively; p values $< .002$). Children scored 6.33 points higher on the 11-item total and 5.64 points higher on the 8-item version. The group differences were mainly the result of significantly higher scores by the children on three items: increased motor activity/energy (item 2), ele-

TABLE 4
P-YMRS (11- and 8-Item Versions) Mean Scores by Diagnostic Categories

Group	n	11-Item		8-Item	
		Mean	SD	Mean	SD
Bipolar I	39	25.40 ^a	8.88	21.82	7.65 ^a
Other bipolar	19	20.21 ^{a,b}	7.19	17.13	7.23 ^{a,b}
Disruptive behaviors	18	13.44 ^{b,c}	7.17	11.56	6.08 ^{b,c}
Unipolar mood	31	11.11 ^c	7.74	8.69	6.19 ^c
Residual	10	10.75 ^c	8.57	8.20	6.71 ^c
Total	117	17.68	10.11	14.84	8.88

Note: P-YMRS = parent version of the Young Mania Rating Scale. ^{a,b,c} Superscripts indicate homogenous subsets according to the Tukeys honestly significant difference test; post hoc p values $< .05$.

TABLE 5
Logistic Regression Analyses Examining Diagnostic Distinctions

Comparison	Scale	Model Fit		Overall Classification ^a (%)	Regression Weights	
		χ^2_1	R^2		MRS	Constant
(a) Bipolar ($n = 58$) vs. All others ($n = 58$)	11	49.8***	.46	77.8	0.18***	-3.08***
	8	51.96***	.48	77.8	0.20***	-2.95***
(b) Bipolar ($n = 58$) vs. Unipolar ($n = 31$)	11	38.5***	.48	79.8	0.19***	-2.61***
	8	40.1***	.50	78.7	0.21***	-2.34***
(c) Bipolar ($n = 58$) vs. ADHD ($n = 18$)	11	19.7***	.34	82.9	0.16***	-1.81***
	8	17.4***	.31	78.9	0.16***	-1.40***

Note: MRS = Mania Rating Scale; ADHD = attention-deficit/hyperactivity disorder.

^a Overall accuracy of classification, also known as total predictive value (e.g., Biederman et al., 1995).

*** $p < .0005$, all two-tailed.

vated mood (item 1), and rate and amount of speech (item 6) (all t_{115} values > 2.86 , p values $< .005$). The higher scores by children on items relating to motor activity, elevated mood, and rate of speech may reflect normal developmental differences in these aspects of behavior. Alternatively, they might be related to symptoms of ADHD, because ADHD was significantly more likely to be diagnosed in the younger cohort ($\chi^2_1 = 9.50$, $p = .002$). It is interesting to note that there was no difference in average rate of endorsement for sexual interest ($t = -0.12$, NS).

Differentiating Diagnostic Categories. Logistic regression analyses tested the extent to which the P-YMRS could differentiate between clinical diagnostic categories based on structured interviews of the child and parent. We tested how well the P-YMRS discriminated between (1) bipolar spectrum disorder versus no diagnosis, (2) unipolar versus bipolar spectrum depression, and (3) bipolar spectrum versus disruptive behavior disorders. The first comparison assessed the performance of the P-YMRS as a screening instrument to detect bipolar spectrum disorders within a sample of nonclinical and mood-disordered individuals. All of the "no diagnosis" participants in the sample presented with enough impairment in functioning that a family member had initiated a referral. Thus the comparison group was likely to show somewhat more pathology than would be expected in a community or epidemiological sample. This could potentially lessen group differences in P-YMRS scale scores, decreasing the likelihood of finding statistically significant results. We included youths with disruptive behavior disorders in comparison 3. This analysis was likely to yield smaller effect sizes than would a comparison between mood disorders versus no diagnosis; however, this analysis more closely paralleled the way clinicians would use the measure for general screening.

Table 5 presents the results of the logistic regression analyses, which predict the odds that a given individual would receive a diagnosis of bipolar spectrum disorder (Biederman et al., 1995; see Hosmer and Lemeshow, 1989, for detailed discussion). As with any regression procedure, these weights are optimized to fit this particular set of data and will be less accurate when applied to new samples. In all comparisons, the P-YMRS resulted in substantial statistical improvement in classification: all χ^2 significant at $p < .0005$, with Nagelkerke R^2 estimates ranging from 0.31 (any bipolar versus disruptive behavior) to 0.50 (any bipolar versus unipolar). In addition, the P-YMRS accomplished respectable classification rates, ranging from 78% to 79% for the 8-item total and 78% to 83% for the 11-item version. The P-YMRS does not appear as successful at differentiating bipolar I from other bipolar disorders, achieving a modest 67% classification rate with both the 8- and 11-item formats.

Diagnostic Efficiency Statistics. ROC analyses determined the relative value of the P-YMRS in making the same diagnostic distinctions discussed above (Chen et al., 1994; McFall and Treat, 1999; Swets, 1992). ROC plots sensitivity and specificity of a diagnostic test while systematically moving the cut score across its full range of values. The accuracy of the ROC can be quantified by calculating the area under its curve (AUC), with chance diagnostic performance corresponding to an AUC of 0.50 and 1.0 indicating perfect performance. The AUC is the probability of the test correctly ranking two randomly selected children into their appropriate diagnostic groups. According to Swets (1992), AUCs between 0.50 and 0.70 show low accuracy, a range of 0.70 to 0.90 represents medium accuracy, and a range of 0.90 to 1.00 denotes high accuracy.

Table 6 presents the AUCs for the ROC analyses of all three comparisons. The 8-item P-YMRS performed fairly

TABLE 6
Areas Under the Curve (AUCs) for Receiver Operating
Characteristic Analyses of P-YMRS

Comparison	8-Item Scale AUC	11-Item Scale AUC
(1) Bipolar ($n = 58$) vs. All others ($n = 58$)	0.86***	0.85***
(2) Bipolar ($n = 58$) vs. Unipolar ($n = 31$)	0.87***	0.87***
(3) Bipolar ($n = 58$) vs. ADHD ($n = 18$)	0.82***	0.82***

Note: AUCs are based on nonparametric estimates, plotting sensitivity as a function of $(1 - \text{specificity})$. P-YMRS = parent version of the Young Mania Rating Scale; ADHD = attention deficit/hyperactivity disorder.

*** $p < .0005$.

well in distinguishing those with bipolar spectrum disorders from those with any other Axis I diagnosis (AUC = 0.86 and 0.85 for the 8- and 11-item versions, respectively). This means that a youth with a diagnosed bipolar spectrum disorder scored higher than a youth with any other Axis I disorder 86% of the time on the P-YMRS. The P-YMRS also proved useful in discriminating youths with bipolar spectrum disorders from those with disruptive behavior disorders (AUC = 0.82 for this clinically difficult distinction).

DISCUSSION

Major Findings

The P-YMRS is established as a one-dimensional scale, indicating that the interpretation of its results best uses the total score. The P-YMRS demonstrates acceptable internal consistency, with an 8-item version appearing better psychometrically in our particular sample.

The P-YMRS discriminated well between youths with formally diagnosed Axis I bipolar spectrum disorders versus those with other Axis I diagnoses, including ADHD or unipolar mood disorders. Group differences between the bipolar spectrum versus the other diagnoses as a set averaged 10.8 points on the 8-item P-YMRS (Cohen $d = 1.71$, using the comparison group's standard deviation, which is more than double the benchmark of $d = 0.80$ for a "large effect size" for the social sciences) (Cohen, 1988). Classification rates were roughly 80% accurate, with diagnostic efficiency statistics such as AUC ranging from 0.82 to 0.87 for specific differential comparisons.

Most importantly, caregivers completing the P-YMRS were able to provide useful information toward making accurate distinctions between youths with bipolar mood disorders versus youths with other diagnoses. The 8-item scale enabled fairly accurate discernment of bipolar spec-

trum disorders versus unipolar mood disorders (79% accuracy, AUC = 0.87) and bipolar spectrum versus disruptive behavior disorders (79% accuracy, AUC = 0.82). This suggests that the P-YMRS may be helpful in making clinically challenging diagnostic distinctions. In our sample, the P-YMRS performed fairly well at separating youths with bipolar spectrum disorders from youths with ADHD. These findings are noteworthy in light of the controversy surrounding diagnosis of bipolar disorder in youths (Biederman et al., 1995, 1998; Carlson, 1998; Geller et al., 1998a,b).

The classification rates based on the P-YMRS are comparable with those reported by other investigators using the parent-reported Child Behavior Checklist (CBCL) to discriminate between ADHD and juvenile mania (Biederman et al., 1995; Geller et al., 1998a). However, the P-YMRS results possess three important advantages. First, the P-YMRS is a shorter instrument than the CBCL (8 or 11 items versus 118). Second, the classifications need only one scale, whereas the CBCL analyses use five to nine syndrome scales. Third, the P-YMRS is based on specific symptom assessment of bipolar psychopathology. The CBCL was not intended to be a measure of bipolar spectrum symptoms; it is unclear whether the discriminating value of high scores on CBCL syndrome scales is due to bipolar individuals showing more severe pathology (Biederman et al., 1995).

The P-YMRS is also a good candidate measure for monitoring response to treatment. The standard error of measurement and standard error of the difference are relatively small, especially when compared with the separations observed between average scores for youths with bipolar spectrum disorders versus other clinical pathology. This means that the P-YMRS combines good differentiation between diagnostic groups with relatively precise assessments of current symptoms.

Strengths of this study include that all diagnoses were made with structured interviews completed by trained research assistants with good overall agreement. Diagnoses were made blind to P-YMRS ratings and were based on youth and parent interviews. Thus there was substantial independence in both the source and the method (questionnaire versus structured interview) of information-gathering. Finally, the study was based on a relatively large sample size.

Limitations

Almost all of the youths participating in the study had presented at an outpatient clinic for evaluation of clini-

cal concerns. However, the elevation of their scores made statistical analyses more conservative, making it less likely that we would find between-group differences. Elevated scores also may increase generalization of these findings to clinical settings, where youths also typically present with psychiatric symptoms.

This study relied on parent report, without similar measures from youths, teachers, or staff. Parents appear to be valid but imperfect informants about their child's cognitive ability (Waschbusch et al., 2000), behavior problems (Richters, 1992; Youngstrom et al., 2000), and emotions (Youngstrom et al., 1999). Parent report on the P-YMRS correlated $r = 0.65$ ($p < .00005$) with the trained clinician ratings on the Y-MRS. A future publication will address comparison of the parent version and clinician completing the original observer-rated Y-MRS in more detail (Youngstrom et al., in press).

This study was performed in a specialized research sample, and our results may not generalize to a clinical setting. In addition, our subject population was predominantly white, limiting generalization of our findings to minority youths. The Y-MRS has been used chiefly as a symptom severity scale after the diagnosis is made.

Clinical Implications

This study requires replication in a large clinical population (which would have a lower incidence of bipolar disorder) before its usefulness in assisting in the diagnostic process in a general clinical setting can be ascertained. The P-YMRS also has not been assessed to determine its ability to detect improvement over time. Its most immediate usefulness may be in epidemiological studies, to explore its performance in screening for bipolar disorder in children and adolescents in community samples.

Future research should also address what additional alternative measures exist for assessing mania and hypomania and identifying optimal strategies for combining parent report with other information sources to maximize diagnostic accuracy. It is important to cross-validate results, to determine potentially robust cutoff scores for clinical and research purposes, and to determine specific item usefulness in another subject sample. This sample did not find items involving sexual interest, thought content, or insight useful. This could represent a regional difference in parents' awareness of or reporting of this material, or an effect of the relatively young mean age group or their outpatient status.

The P-YMRS appears to be a promising adaptation of an established instrument that may contribute substantially both to research and clinical work with bipolar spectrum disorders in youths.

REFERENCES

- Baldessarini RJ, Finklestein S, Arana GW (1983), The predictive power of diagnostic tests and the effect of prevalence of illness. *Arch Gen Psychiatry* 40:569–573
- Biederman J, Russell R, Soriano J, Wozniak J, Faraone SV (1998), Clinical features of children with both ADHD and mania: does ascertainment source make a difference? *J Affect Disord* 51:101–112
- 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
- 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
- Chen WJ, Faraone SV, Biederman J, Tsuang MT (1994), Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *J Consult Clin Psychol* 62:1017–1025
- Cohen J (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum
- Fristad MA, Weller EB, Weller RA (1992a), Bipolar disorder in children and adolescents. *Child Adolesc Psychiatr Clin N Am* 1:13–29
- Fristad MA, Weller EB, Weller RA (1992b), The Mania Rating Scale: can it be used in children? A preliminary report. *J Am Acad Child Adolesc Psychiatry* 31:252–257
- Fristad MA, Weller RA, Weller EB (1995), The Mania Rating Scale (MRS): further reliability and validity studies with children. *Ann Clin Psychiatry* 7:127–32
- Geller B, Warner K, Williams M, Zimmerman B (1998a), 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, Williams M, Zimmerman B, Frazier J, Beringer L, Warner KL (1998b), Prepubertal and early adolescent bipolarity differentiate from ADHD by manic symptoms, grandiose delusions, ultra-rapid or ultradian cycling. *J Affect Disord* 51:81–91
- Gracious BL, Holmes WD, Ruppert N, Burke KC, Hurt J (1994) Mania Rating Scale reliability in children and adolescents. Presented at the First Annual International Conference on Bipolar Disorders, Pittsburgh, June
- Hatterer JA, Kocsis JH, Stokes PE (1988), Thyroid function in patients maintained on lithium. *Psychiatry Res* 26:249–57
- Hosmer DW, Lemeshow S (1989), *Applied Logistic Regression*. New York: Wiley
- Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH (1994), *Biostatistics in Clinical Medicine*. New York: McGraw-Hill
- Jacobson NS, Truax P (1991), Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 59:12–19
- 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
- Kraemer HC (1992), *Evaluating Medical Tests: Objective and Quantitative Guidelines*. Newbury Park, CA: Sage
- McFall RM, Treat TA (1999), Quantifying the information value of clinical assessment with signal detection theory. *Annu Rev Psychol* 50:215–241
- Poolsup N, Li Wan Po A, Oyeboode F (1999), Measuring mania and critical appraisal of rating scales. *J Clin Pharm Ther* 24:433–443
- Richters JE (1992), Depressed mothers as informants about their children: a critical review of the evidence for distortion. *Psychol Bull* 112:485–499
- Streiner DL, Norman GR (1995), *Health Measurement Scales: A Practical Guide to their Development and Use*, 2nd ed. New York: Oxford University Press

- Swets JA (1992), The science of choosing the right decision threshold in high-stakes diagnostics. *Am Psychol* 47:522–532
- Velicer WF, Eaton CA, Fava JL (2000), Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components. In: *Problems and Solutions in Human Assessment: A Festschrift to Douglas Jackson at Seventy*, Goffin RD, Helmes E, eds. Hillsdale, NJ: Erlbaum, pp 41–71
- Waschbusch DA, Daleiden E, Drabman RS (2000), Are parents accurate reporters of their child's cognitive abilities? *J Psychopathol Behav Assess* 22:61–77
- Weller RA, Weller EB, Tucker SG, Fristad MA (1986), Mania in prepubertal children: has it been underdiagnosed? *J Affect Disord* 11:151–154
- Wiggins JS (1973), *Personality and Prediction: Principles of Personality Assessment*. Reading, MA: Addison-Wesley
- 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, Ackerman BP, Izard CE (1999), Dysphoria-related bias in maternal ratings of children. *J Consult Clin Psychol* 67:905–916
- Youngstrom EA, Gracious BL, Danielson CK, Findling RL, Calabrese JR (in press), Toward an integration of parent and clinician report on the Young Mania Rating Scale. *J Affect Disord*
- Youngstrom EA, Loeber R, Stouthamer-Loeber M (2000), Patterns and correlates of agreement between parent, teacher, and male youth behavior ratings. *J Consult Clin Psychol* 68:1038–1052