Published in final edited form as:

J Clin Child Adolesc Psychol. 2015 ; 44(6): 992–1007. doi:10.1080/15374416.2014.915548.

Unfiltered Administration of the YMRS and CDRS-R in a Clinical Sample of Children

Andrea M. Yee, **Guillermo Perez Algorta**, **Eric A. Youngstrom**, **Robert L. Findling**, **Boris Birmaher**, **Mary A. Fristad**, and **The LAMS Group**

The Ohio State University

Andrea M. Yee: yee.115@osu.edu; Guillermo Perez Algorta: guillermo.perezalgorta@osumc.edu; Eric A. Youngstrom: eay@unc.edu; Robert L. Findling: RFindli1@jhmi.edu; Boris Birmaher: BirmaherB@MSX.UPMC.EDU; Mary A. Fristad: mary.fristad@osumc.edu

Abstract

Objective—The objective of this study is to evaluate discriminative validity of the Young Mania Rating Scale (YMRS) and Children's Depression Rating Scale – Revised (CDRS-R) in a clinical sample of children when administered in an unfiltered manner (i.e., regardless of whether symptoms occur in a mood episode).

Method—The Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) is the gold standard for assessing psychiatric disorders in children and was used to make diagnoses in this study. Using a sample of 707 treatment-seeking youth (aged 6–12 years, M_{age} = 9.7 years, 67.6% male), receiver operating curve analyses were performed and diagnostic likelihood ratios (DLRs) were calculated to evaluate the ability to change the odds and differentiate bipolar disorder (BD) from other disorders (using the YMRS) and depression from other disorders (using the CDRS-R).

Results—Using unfiltered administration, the YMRS achieved good discriminative validity when classifying BD compared to other disorders (AUC = .86) and increased odds of a bipolar diagnosis given a score in the highest quintile (DLR = 6.12). Using unfiltered administration, the CDRS-R achieved moderate to good discriminative validity in classifying depressive disorders (DD) compared to other disorders ($AUC_{BD\ in\ comparison}$ = .78; $AUC_{BD\ not\ in\ comparison}$ = .84) and slightly increased odds of DD given a score in the highest quintile ($DLR_{BD\ in\ comparison}$ = 3.12; $DLR_{BD\ not\ in\ comparison}$ = 5.08).

Conclusions—The YMRS and CDRS-R have moderate to good discriminative validity when administered in an unfiltered way in a sample of treatment seeking youth.

Keywords

bipolar disorder; children; Young Mania Rating Scale (YMRS); Children's Depression Rating Scale; Revised (CDRS-R); discriminative validity

Correspondence concerning this article should be addressed to Andrea M. Yee (Attn: Dr. Mary A. Fristad), Department of Psychiatry, The Ohio State University, 1670 Upham Drive, Suite 460G, Columbus, OH 43210. yee.115@osu.edu.

Andrea M. Yee, Department of Psychology, The Ohio State University, 1670 Upham Drive, Suite 460G, Columbus, OH 43210; Guillermo Perez Algorta, Department of Psychiatry, The Ohio State University, Columbus, OH; Eric A. Youngstrom, Departments of Psychology and Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC; Robert L. Findling, Department of Psychiatry, John Hopkins University, Baltimore, MD; Boris Birmaher, Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA; Mary A. Fristad, Departments of Psychiatry, Psychology and Nutrition, The Ohio State University, 1670 Upham Dr, Columbus, OH 43210-1250.

Pediatric bipolar disorder (BD) is associated with a host of negative outcomes, including substance abuse, school failure, aggression, and suicide (Lewinsohn, Klein, & Seeley, 1995; Findling et al., 2001; Geller et al., 2003; Lewinsohn, Seeley, & Klein, 2003). Given its early onset and chronicity, persons with BD are more likely than the general population to have illnesses such as obesity, diabetes, and heart disease (Goldstein, Liu, Schaffer, Sala, & Blanco, 2013), develop them earlier, and die from them sooner (World Health Organization, 2011). In adults, the aggregate cross-study estimated lifetime prevalence of BD is about 1%, with a range of 0.8% to 3.3% (Merikangas & Pato, 2009). In youth, disparate rates of BD are reported in epidemiological studies, with hypomania ranging from 0% to 0.9% and mania ranging from 0.4% to 1.9% (Merikangas & Pato, 2009). A meta-analysis of published epidemiologic studies found the overall rate of pediatric bipolar disorder to be 1.8% (95% CI, 1.1%–3.0%; Van Meter, Moreira, & Youngstrom, 2011).

There has been a dramatic increase in the diagnosis of BD in youth since the mid-1990s (Danner et al. 2009; Moreno et al. 2007; Blader & Carlson, 2007). Rates at hospital discharge increased linearly from 1.3 to 7.3 per 10,000 from 1996 to 2004 (Blader & Carlson, 2007). In outpatient clinics, rates range from 6% to 17%; one inpatient study found 30% of youth had manic symptoms (Youngstrom & Duax, 2005). As there is no secular trend in epidemiological rates for community samples (Van Meter et al., 2011), these increased clinical rates most likely reflect changes in awareness and clinical practice. Additionally, BD in youth has received increased scientific attention; the number of research articles increased from less than 500 between 1986–1990 to over 2000 between 2006–2010 (Fristad & Algorta, 2013, p. 738).

As research on BD in children continues to increase, it is important to examine the most commonly used outcome measures. The quality, applicability, and utility of clinical trials of pediatric BD rely on their accurate and appropriate interpretation. The American Academy of Child and Adolescent Psychiatry, as part of their Best Practice efforts, recommends using the Young Mania Rating Scale (YMRS; Young et al., 1978) and the Children's Depression Rating Scale – Revised (CDRS-R; Poznanski et al., 1984) (Carlson et al., 2003); they are the two most commonly used outcome measures in pediatric BD research (Youngstrom, Findling, Youngstrom, & Calabrese, 2005).

Both the YMRS and CDRS-R assess symptom severity. They are not intended to be diagnostic tools nor do they cover all diagnostic criteria. They can be administered two ways: "filtered" or "unfiltered". Filtered measures refer to assessments that take into account whether or not symptoms occur within the context of a mood episode. They tend to be lengthy semi-structured interviews, taking into account onset, duration, lifetime occurrence, baseline functioning and changes from it, as well as symptom episodicity and chronicity. Current symptoms are interpreted differentially (e.g., chronic hyperactivity consistent with a child's attention deficit hyperactivity disorder [ADHD] would not be counted as a symptom whereas episodic [or significantly exacerbated] hyperactivity that occurs in the context of other mood symptoms would be counted). Therefore, filtered measures incorporate information that would help with diagnostic formulation. Symptoms counted during filtered

administration of the YMRS and CDRS-R are not manifestations of other childhood disorders.

Unfiltered measures tend to be clinician-rated or self-report measures that are fairly quick to complete. They capture current symptom presentation and severity regardless of change from baseline functioning and without considering the context of mood episodes. They do not differentiate chronic from episodic symptoms, nor do they account for symptom onset or duration. Therefore, unfiltered measures can be quick to administer and are intended to capture symptom severity but may be less suited for use as diagnostic tools.

YMRS and CDRS-R administration is considered unfiltered if conducted in a "what you see is what you get" manner. Therefore, symptoms counted on the YMRS and CDRS-R, when administered in an unfiltered way, may be manifestations of other childhood disorders that have symptom overlap with mood disorders (see Tables 1–2).

The YMRS is an 11-item clinician-rated scale originally designed for use with inpatient adults, where it showed good reliability and validity. It was created to measure symptom severity after a manic episode had been determined. Item scores were based on the past 48 hours and clinician observation during the interview, with an emphasis on the latter (Young et al, 1978), and was intended to get a snapshot of patients' manic symptom severity. In youth aged 6–12, the YMRS has been found to differentiate inpatient children with BD from inpatient and outpatient children with ADHD (Fristad, Weller, & Weller, 1992; Fristad, Weller, & Weller, 1995). Frazier et al. (2007) found efficiency of the YMRS for discriminating BD from other disorders was excellent (AUCs = 0.92–0.99) in children aged 4–17 who were divided into four age groups: 4–7 years, 8–10 years, 11–13 years, and 14–17 years. These studies explicitly stated the YMRS was administered in a filtered way: "Scores were made by the clinician after combining impressions from the child's and parents' clinical interviews," or the YMRS was completed by a clinician who had first administered a detailed interview, indicating the YMRS was administered in a filtered way (Fristad, Weller, Weller, 1992 p. 253; Frazier et al., 2007). Another study found the YMRS was able to discriminate children aged 8–17 with ADHD and mania from those with ADHD without mania, but the YMRS method of administration was not reported (Serrano, Ezpeleta, Alda, Matali, & San, 2011).

The CDRS-R is a 17-item clinician-administered and rated scale, with 14 items scored from verbal responses and 3 items scored from observation during the interview (Poznanski et al. 1984). It is intended to be administered as a semi-structured interview as a filtered measure (Poznanski, Freeman, & Mokros, 1985). Children with depression have been found to receive significantly higher mean total scores than children without depression (Poznanzki et al., 1984).

Therefore, previous studies have demonstrated the YMRS and CDRS-R are able to discriminate BD and depression, respectively, from other disorders or no disorder when administered in a filtered way (Poznanzki et al., 1984; Fristad, Weller, & Weller, 1992; Fristad, Weller, & Weller, 1995; Frazier et al., 2007).

Clinical research on pediatric BD uses the YMRS and CDRS-R as outcome measures with the assumption that they are measuring manic and depressive symptomatology, respectively. Previous studies that have used filtered administration support this claim. However, unfiltered administration has not been explicitly studied. Since administration method of the YMRS and CDRS-R is rarely reported in clinical research, it is important to examine the discriminative validity of these two commonly used outcome measures when administered in an unfiltered manner.

Unfiltered administration of the YMRS and CDRS-R may be the preferred method because it is faster. Additionally, unfiltered administration may be the default method in clinical or research settings where clinicians are not trained or instructed to clearly determine the presence of a mood episode before administering the YMRS and CDRS-R. In settings that emphasize symptom presentation, unfiltered administration may be preferred and more practical.

For two main reasons, we expected unfiltered administration of the YMRS and CDRS-R to show significantly lower discriminative validity than previous findings based on filtered administration (e.g., Youngstrom et al., 2005; Frazier et al., 2007). First, there is substantial nosological overlap between mood disorders (i.e., mania and depression) and other more common childhood disorders such as ADHD, disruptive behavior disorders (i.e., Oppositional Defiant Disorder [ODD]/Conduct Disorder [CD]), and anxiety disorders (Tables 1–2). Thus, the same symptom could occur due to a variety of diagnoses or developmental pathways. Second, pediatric BD has high rates of comorbidity. ADHD, disruptive behavior disorders (i.e., ODD/CD), and anxiety disorders occur along with BD at higher rates than expected in the general population. ADHD has been found to be comorbid with BD 59% – 93% of the time (Carlson & Meyer, 2006); a meta-analysis reported 67% (Kowatch, Youngstrom, Danielyan, & Findling, 2005). ODD/CD is comorbid with BD 43 – 91% of the time and anxiety disorders 12 – 59% of the time (Carlson & Meyer, 2006). These high comorbidity rates complicate assessment of mania because “textbook” presentations of mania are rare and common comorbid conditions have higher base rates than BDs (Youngstrom, Findling, Youngstrom, & Calabrese, 2005).

There are serious methodological and clinical implications if unfiltered administration of the YMRS and CDRS-R has poor discriminative validity. If so, the quicker, unfiltered administration of the YMRS and CDRS-R would not be able to reliably differentiate manic and depressive symptoms from other symptomatology, invalidating statements that they measure specifically manic and depressive symptoms and compromising their efficacy as mood symptomatology outcome measures in clinical research.

The purpose of this study was to examine discriminative validity of the YMRS and CDRS-R when administered in an unfiltered way. Our first hypothesis was that total scores of unfiltered YMRS and CDRS-R would be significantly higher among cases with BD and depressive disorders (DD), respectively. Our second set of hypotheses was that unfiltered administration would result in significantly lower discriminative validity than previously found with filtered administration of the same instrument for the same diagnostic comparisons. We predicted that unfiltered administration of the YMRS and CDRS-R would

have significantly lower discriminative validity due to nosological overlap and comorbid presentations than reported in prior work using filtered ratings (e.g., Youngstrom et al., 2005; Frazier et al., 2007).

Method

Participants

The LAMS study is a multicenter study conducted at nine outpatient clinics associated with four locations: Case Western Reserve University, Western Psychiatric Institute and Clinic, Cincinnati Children's Hospital Medical Center, and The Ohio State University. The LAMS study is a prospective, longitudinal study of 621 children with elevated symptoms of mania (ESM+) and a comparison group of 86 children without elevated symptoms of mania (ESM −). A thorough description of recruitment strategies, sample characteristics and methodologic detail can be found in prior publications (Horwitz et al., 2010; Findling et al., 2010).

The current study analyzed baseline data from 707 LAMS children. Children's ages ranged from 6 years 0 months to 12 years 11 months of age (average, 9.7 years) A majority was male (68%), white (64%), and receiving Medicaid (52%); a minority lived with both parents (32%). Approximately one-quarter (n=162, 23%) had a bipolar spectrum disorder (BPSD): 77 had bipolar disorder NOS (BP-NOS), 71 had bipolar I disorder (BP-I), 11 had cyclothymic disorder, and 3 had bipolar II disorder (BP-II).

Procedures

The study was approved by institutional review boards at each university-affiliated LAMS site. Informed consent was obtained before screening; informed consent and assent were obtained before baseline, prior to any respective study procedures. Screening was conducted using 10 items from the 73-item Parent General Behavior Inventory (PGBI) that best discriminate BD from other disorders; the PGBI-10M has been found to have excellent reliability (Findling et al., 2002; Youngstrom et al., 2004; Youngstrom, Findling, Danielson, & Calabrese, 2001; Youngstrom, Frazier, Demeter, Calabrese, & Findling, 2008). All children with PGBI-10M scores ≥ 12 (i.e., ESM+) were invited to participate. One ESM− (i.e., score ≤ 11) participant similar in age (within 2 years), sex, race/ethnicity, and type of insurance as the "modal" ESM+ participant, was invited to enroll in the longitudinal portion of the study for every ten consecutive ESM+ participants enrolled. Groups did not significantly differ on any previously mentioned variables (Horwitz et al., 2010; Findling et al., 2010). Baseline assessment was conducted 3–6 weeks after screening. Participants were compensated $80 for a baseline assessment that could last 7–8 hours (based on a pay scale of $10/hour). Information gathered at baseline included: 1) demographics; 2) functional assessment; 3) diagnoses; and 4) symptomatic assessment.

Interviewers

Baseline assessments were conducted by interviewers, with an educational background that ranged from post-baccalaureate to post-doctoral, who had experience with psychiatrically impaired children and completed extensive training with strict inter-rater reliability

requirements. Interviewers were required to match seven diagnostic categories on the K-SADS-PL-W, and obtain satisfactory item-level weighted kappas ($\kappa \geq .40$) for the CDRS-R, YMRS, and KSADS-PL-W. Item-level kappas are partly dependent on the size of the item pool, similar to Cronbach's alpha, so the same threshold is harder to achieve with shorter scales; however, values >.40 are considered adequate in all conventional benchmarking schemes. Raters achieved a κ of 0.82 for all KSADS-PL-W psychiatric diagnoses, 0.93 for bipolar diagnoses, 0.47 for the CDRS-R, and 0.41 for the YMRS. Note that kappas at the item level are different in structure, calculation, and typical benchmarks than kappas at the diagnosis level (Siegel & Castellan, 1988). The same interviewer conducted the YMRS, CDRS-R, and the K-SADS-PL-W (Kaufman et al., 1997; Geller et al., 1998) with the parent and child. Interviewers were trained to administer the YMRS and CDRS-R in an unfiltered manner. The YMRS and CDRS-R were administered before the KSADS-PL-W, ensuring that the interviewer was blind to diagnosis at time of administration.

Measures

Diagnoses were made with the K-SADS-PL-W, the gold standard for assessing psychiatric disorders in children. Many versions have been adapted from the original KSADS; all are semi-structured interviews with demonstrated good reliability and validity (Chambers et al, 1985; Kaufman et al., 1997; Geller et al., 1998; Geller et al., 2001). The K-SADS-PL-W ascertained presence of manic and depressive symptoms within the context of a mood episode (i.e., filtered rating).

The YMRS and CDRS-R provided unfiltered ratings of symptom severity. The YMRS includes 11-items; seven are rated from 0 (Absent) to 4; four from 0 to 8; total scores range from 0 to 60. Scores were obtained via separate interviews with both parent and child regarding symptoms occurring over the past 2 weeks. Interviewers used clinical judgment to resolve discrepancies between parent and youth report. Cronbach's alpha for the YMRS in this sample was .76.

The CDRS-R is a 17-item scale administered to both parent and child in separate interviews about symptoms over the past 2 weeks. Parents were asked all items. Children were asked 14 items and were rated based on observation for 3 items: Depressed Affect, Tempo of Speech, and Hypoactivity. Total scores range from 17 (no symptoms) to 113. Three items are rated from 1 (No Symptom) to 5; 12 items are rated from 1 to 7. Cronbach's alpha for the CDRS-R in the current sample was .81.

Administration of the YMRS and CDRS-R was explicitly administered before the KSADS-PL-W in an unfiltered way:

> "These unfiltered ratings did not require clinical judgment about the reasons for symptoms to be manifest. Because a key aspect of the LAMS study is the assessment of symptoms, regardless of etiology, over time, these unfiltered ratings were obtained to complement those assessments of affective illness that were manifest only during the presence or a mood disorder,"
>
> (Findling et al., 2010 p. 1666).

Diagnosis of Bipolar Spectrum Disorders

Diagnoses of BP-I, BP-II, and cyclothymic disorder used unmodified DSM-IV criteria. We operationalized BP-NOS as: 1) elated mood plus 2 associated symptoms of mania or irritable mood plus 3 associated symptoms of mania; 2) change in functioning level; 3) symptom duration must be ≥ 4 hours within a 24-hour period; and 4) four episodes of ≥ 4 hours over 4 days must be present in the participant's lifetime, following the Course of Bipolar Youth (COBY) study criteria for BP-NOS (Axelson et al., 2006).

Data Analytic Strategies

YMRS analyses compared children with BD (n=162) to those without BD (n=545). Additional analyses compared BD to specific diagnostic groups to address the two factors most likely to influence discriminative validity of unfiltered administration: comorbidity and symptom overlap (see Table 1). Diagnostic comparison groups included: 1) children with ODD/CD (n=293) who had any comorbidity, ("complicated ODD/CD group"); 2) children with autism spectrum disorders (ASD (n=32) who had any comorbidity except ODD/CD; 3) children with ADHD (n=157) who did not have comorbid ODD/CD or ASD but could have other comorbidities ("uncomplicated ADHD group"); and 4) children with anxiety (ANX) (PTSD, acute stress disorder, OCD, panic disorder, separation anxiety disorder, specific phobia, social phobia, GAD, and/or anxiety disorder NOS) (n=49) who did not have comorbid ODD/CD, ASD, or ADHD, but could have other comorbidities.

CDRS-R analyses used three different stratifications. First, to increase external validity, children with DD (n=124) were compared to all others, including those with BD in comparison groups (n=583). Second, to address the complication that children with BD experience depressive symptoms, children with BD were excluded, creating a "pure' comparison of DD versus non-mood disorders (n=421). Third, to address the clinical reality that the CDRS-R is used to assess both BD and DD, a mood group was formed by combining BD and DD (n=286). All CDRS-R analyses were conducted comparing those with DD, or a mood disorder in the third stratification, to those without DD, or any mood disorders in the third stratification. Additional comparisons addressed symptom overlap and comorbidity, which are likely to impact the discriminative validity of unfiltered CDRS-R administration (see Table 2). Comparison groups consist of : 1) children with ODD/CD and any comorbidity; and 2) children with ANX and any comorbidity except ODD/CD (see Table 6).

Receiver Operating Characteristic (ROC) analyses examined how well unfiltered YMRS and CDRS-R administration predicted presence or absence of BD or DD, respectively, in this sample. ROC was originally used as an electrical signal detection procedure in World War II to detect signal from noise (Metz, 1986). As the most basic property of a test of diagnostic accuracy, discriminative validity should reflect a test's ability to distinguish between two states of health (e.g., disease versus no disease). The most common index of accuracy in ROC analysis is area under the curve (AUC), the probability of correctly distinguishing two randomly chosen participants. An AUC of .50 means the test does not distinguish more than chance, whereas an AUC of 1.0 means the test distinguishes perfectly. Common standards for interpreting AUC values are: .60–.70, poor accuracy; .70–.80,

moderate accuracy; .80–.90, good accuracy; and .90–1.00, high accuracy (Swets, 1988). However, in mental health applications, good tests often provide AUC values of .70 to .80; values higher than .90 frequently indicate flawed designs in which the comparison group consists of healthy controls or other comparators of minimal clinical relevance (Bossuyt et al., 2003; Youngstrom, Meyers, Youngstrom, Calabrese, & Findling, 2006). We tested whether the YMRS performed significantly less well when administered in an unfiltered manner by testing the difference between the AUC from the present sample versus the filtered administration results in separate data published by Youngstrom et al. (2005) and Frazier et al. (2007) using Hanley and McNeil's (1983) z-test of independent AUCs. The same method was used for the CDRS-R.

We estimated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for cut points that maximized the combination of sensitivity and specificity, treating costs of false positive and false negatives as equal. We estimated multi-level diagnostic likelihood ratios (DLRs) based on YMRS and CDRS-R quintiles. DLRs are the change in odds of a given outcome (e.g., diagnosis of BD) and can be derived from sensitivity and specificity (Straus, Glasziou, Richardson, & Haynes, 2011). DLRs can be used to find posterior probability, which is the new probability of a disorder given a certain test result. A visual nomogram, analogous to a slide rule, is used to combine previous probability (i.e., base rate) with the test result (i.e., the DLR corresponding to the appropriate quintile) to find the posterior probability, which is equivalent to PPV. Nomograms use nonlinear spacing to accomplish these transformations, which would require several computations; it is the same use of geometry that handles multiplication and ratios underlying old-fashioned slide rules (Straus, Glasziou, Richardson, & Haynes, 2011). DLRs greater than 10 or less than 0.1 are considered clinically decisive because they can change a prior probability of 50% to more than 90% or less than 10% posterior probability (Youngstrom, Meyers, Youngstrom, Calabrese, & Findling, 2006a).

Results

Preliminary Analyses

Given the small amount of missing data (< 5%), listwise deletion provided an adequate solution (Allison, 2002). Table 3 presents demographics for participants with BD and comparison groups. Participants with BD were significantly older than all comparison groups except ASD and ANX. Although statistically significant, the absolute age differences were unremarkable (< 1 year) and Cohen's *d* effect sizes were small (largest $d = 0.35$). The BD group contained a significantly lower proportion of males and white participants than all comparison groups except for the ASD and ANX groups. The BD group had significantly lower CGAS scores, indicating poorer overall functioning, than the complicated ODD/CD, uncomplicated ADHD, and ANX groups, but significantly higher CGAS scores than the ASD group. All groups (BD, No BD, ODD/CD, ADHD, and ANX) had mean CGAS scores within the same moderate impairment range, except for the ASD group, which had an average CGAS score in the serious impairment range.

Participants within the DD group and within the mood group (Table 4) were significantly older than all comparison groups. Although statistically significant, the difference between

average ages did not exceed 1.1 years, and Cohen's *d* effect sizes were small to medium (largest *d* = 0.52). The DD and mood groups had significantly lower proportions of white participants. The ANX group was comprised of a significantly lower proportion of participants on Medicaid than the DD or mood groups. The DD and mood groups had significantly lower CGAS scores (i.e., poorer overall functioning) than all comparison groups. Of note, all groups had average CGAS scores within the same moderate impairment range.

DSM-IV Disorders and Comorbidity

Comorbid disorders at listed in Tables 5–6. The most common comorbid diagnosis was ADHD, occurring in 72% of participants with BD, 77% of participants with DD, and 74% of participants with either mood disorder. Within the BD group, 42% had ODD/CD, 40% had ANX, 31% had some other diagnosis (e.g., enuresis), and 3% had ASD. Within the DD group, 71% had ANX, 56% had ODD/CD, 30% had some other diagnosis, and 2% had ASD. Within the mood group, 57% had BD, 54% had ANX, 48% had ODD/CD, 43% had DD, 30% had some other diagnosis, and 2% had ASD.

Discriminative Validity of the Unfiltered YMRS Administration

ROC analyses evaluated discriminative validity of the YMRS. When participants with BD were compared to all others, accuracy of predicting BD was in the good range (AUC = .84, 95% CI = .80–.87). The optimal cut point was 22.5. Sensitivity for the YMRS when BD was compared with all other disorders was .68 and specificity was .86. As hypothesized, unfiltered administration was significantly less discriminating than filtered administration (AUC = .95, SE = 0.013), $z = 7.65$, $p < .0005$ and (AUC = .97, SE = 0.012), $z = 5.89$,$p < .0005$, comparing present results to those for treatment seeking youth aged 5–18 in Youngstrom et al. (2005) and aged 4–14 in Frazier et al. (2007), respectively, using the Hanley & McNeil (1983) procedure.

Predicting BD also fell within the good range when compared to the complicated ODD/CD group (AUC = .80, 95% CI = .76–.84); the ASD group (AUC = .80, 95% CI = .72–.88); and the uncomplicated ADHD group (AUC = .88, 95% CI = .84–.92). High accuracy was achieved when discriminating BD from ANX (AUC = .90, 95% CI = .86–.95). Table 7 presents AUCs for all YMRS comparisons, sensitivity, specificity, PPV, and NPV for the observed sample base rates.

Controlling for age, gender, race, and CGAS scores had minimal effects on AUC values for BD versus No BD, ODD/CD, ADHD, and ANX (AUC values changed 0.01–0.03). AUC increased 0.10 for the BD versus ASD comparison when controlling for demographics.

Discriminative Validity of the Unfiltered CDRS-R Administration

ROC analyses also tested discriminative validity of the CDRS-R. Three sets of analyses addressed complications of comparing participants with DD to participants with BD, as both experience depressive symptoms. First, the DD group was compared to all other participants, with BD folded into the comparison group. In these analyses, accuracy of predicting DD was moderate (AUC = .78, 95% CI = .74–.82), with an optimal cut point of

38.5 and corresponding sensitivity of .73 and specificity of .74. When BD was folded into comparison groups, unfiltered administration was not significantly less discriminating than filtered administration (AUC = .81, SE = 0.015), $z = 1.05$, $p = .295$, comparing present results to those for treatment seeking youth aged 5–18 in Youngstrom et al. (2005) using the Hanley & McNeil (1983) procedure.

Controlling for age, gender, race, and CGAS scores had minimal effects on AUC values for DD versus No DD and ODD/CD (AUC values changed 0.02–0.03). AUC increased 0.07 for the DD versus ANX comparison when controlling for demographics.

Second, participants with BD diagnoses were removed from the CDRS-R ROC analysis, leaving a "pure" comparison of DD versus all other non-mood diagnoses. Accuracy improved and was within the good range (AUC = .84, 95% CI = .80–.88), with an optimal cut point of 36.5 and corresponding sensitivity of .76 and specificity of .79. For the pure comparison of DD versus all non-mood diagnoses, unfiltered administration was significantly less discriminating than filtered administration (AUC = .93, SE = 0.010), $z = 3.94$, $p < .0005$, comparing present results to those of Youngstrom et al. (2005) using the Hanley & McNeil (1983) procedure. Controlling for age, gender, race, and CGAS scores had minimal effects on AUC values for DD versus No DD and ODD/CD (AUC values changed 0.02). AUC increased 0.07 for the DD versus ANX comparison when controlling for demographics.

Third, BD and DD were combined into a mood group and compared to all other participants, which had moderate accuracy of distinguishing mood disorders (AUC = .77, 95% CI = .74–.81). Unfiltered administration was significantly less discriminating than filtered administration (AUC = .93, SE = 0.009), $z = 7.65$, $p < .0005$, comparing present results to those of Youngstrom et al. (2005), using the Hanley & McNeil (1983) procedure. Table 7 presents AUCs for all CDRS-R comparisons, sensitivity, specificity, PPV, and NPV for sample base rates in the present sample. Controlling for age, gender, race, and CGAS scores AUC values increased 0.04–0.07 for mood comparisons.

Change in Odds of a BD or DD Diagnosis: Diagnostic Likelihood Ratios (DLRs)

Multi-level DLRs were calculated for YMRS and CDRS-R quintiles to examine change in odds of a BD or DD diagnosis given different test scores. For the YMRS, DLRs were calculated for all comparison groups except ASD. DLRs could not be calculated for all quintiles of the ASD comparison because some cells had zero participants due to the small sample size of the ASD group. Scores in the very high quintile for all remaining comparisons (No BD, complicated ODD/CD, uncomplicated ADHD, and ANX) increased odds of BD and ranged from 5.74 for BD versus complicated ODD/CD to over 10 for BD versus uncomplicated ADHD and BD versus ANX, with the DLR for BD versus all other disorders combined falling in between at 6.12 (Table 8). Scores in the very low range for the YMRS decreased odds of BD and ranged from 0.09 for BD versus uncomplicated ADHD and BD versus ANX to 0.19 for BD versus complicated ODD/CD, with the DLR for BD versus all other disorders combined falling in between at 0.10.

As Figure 1 illustrates, nomograms combine the base rate of BD with DLRs to find the probability of BD given a YMRS score within a certain range. For example, the base rate of BD in this sample was 23%. The DLR corresponding to a YMRS score in the highest quintile was 6.12. Plotted on the nomogram, the odds of a BD diagnosis increases from 23% to just over 60%. Therefore, given a YMRS score in the highest range, the probability of BD is 60%.

DLRs were calculated for three stratifications of DD (Table 9). Including BD in the comparison groups, DLRs associated with the highest CDRS-R quintile were similar and ranged from 3.12 to 3.56. Excluding BD from comparison groups, DLRs associated with the highest CDRS-R quintile ranged from 5.08 for DD versus No DD to 14.37 for DD versus ANX. Combining BD and DD to form a mood group, DLRs associated with the highest CDRS-R quintile ranged from 3.92 for Mood versus ANX to 7.56 for Mood versus ODD/CD.

Overall, the YMRS and CDRS-R, when administered in an unfiltered manner, have moderate to good discriminative validity and are able to increase or decrease the odds of a BD or DD diagnosis given scores within the very high or very low scale ranges.

Discussion

It was hypothesized that unfiltered YMRS and CDRS-R administration would result in significantly lower discriminative validity due to nosological overlap and comorbid presentations than reported in prior work using filtered ratings (e.g., Youngstrom et al., 2005; Frazier et al., 2007). As hypothesized, the YMRS and the CDRS-R, specifically for the pure comparison of DD versus non-mood disorders and the combined BD and DD mood group versus non-mood disorders, showed significantly lower AUCs in ROC analyses than found using filtered scores. Even so, the YMRS still achieved good discriminative validity and the CDRS-R achieved moderate discriminative validity using unfiltered administration (Table 7).

Examining sub-comparisons of BD with other diagnostic groups (complicated ODD/CD, ASD, uncomplicated ADHD, and ANX), the YMRS was best at discriminating BD from uncomplicated ADHD (i.e., no ODD/CD or ASD comorbidities) and BD from ANX. Although the difference between AUCs could not be statistically tested due to different comparison groups, this trend is consistent with expectations. Uncomplicated ADHD and ANX groups lack some of the nosological overlap and much of the qualitative overlap that children with complicated ODD/CD have with BD (e.g., temper outbursts, anger, aggression, and irritability), consistent with previous literature (e.g., Mick, Spencer, Wozniak, & Biederman, 2005).

Looking at CDRS-R sub-comparisons (ODD/CD and ANX), it consistently performed the worst when the DD or mood group was compared to ANX. Although the difference between AUCs could not be statistically tested due to different comparison groups, this trend is consistent with expectations. Depression and anxiety are highly comorbid and have substantial nosological overlap (e.g., difficulty concentrating, insomnia, appetite changes).

In addition, this makes theoretical and empirical sense given biological/neural findings and developmental trajectory studies of internalizing problems (i.e., anxiety in childhood as a pathway to depression in adolescence) (Cummings, Caporino, & Kendall, 2013; Lonigan, Phillips, & Hooe, 2003; Mineka, Watson, & Clark, 1998).

Sensitivity and specificity, which are independent of base rate, were acceptable for both the YMRS and CDRS-R. Positive predictive value (PPV) and negative predictive value (NPV) for the YMRS were good. For the CDRS-R, PPV was poor and NPV was good, indicating the number of false positives exceeded or was almost equivalent to the number of true positives. It is important to note PPV and NPV are affected by base rate. In particular, PPV becomes inflated as base rates increases, and NPV tends to be high when the target condition is rare. For the total sample, the BD base rate was .23, which may be similar to some inpatient settings but is probably higher than outpatient settings and is certainly higher than epidemiologic samples (Soutullo et al., 2005; Youngstrom & Duax, 2005; Danner et al., 2009; Merikangas & Pato, 2009; Van Meter, Moreira, & Youngstrom, 2011). The total sample DD base rate was .18, which may be similar or lower than some outpatient settings but higher than epidemiologic estimates (Costello, Erkanli, & Angold, 2006). Therefore, PPV and NPV are difficult to generalize to other settings with varying base rates. Evidence Based Medicine advocates the probability nomogram, a scale analogous to a slide rule, and other implementations of Bayesian methods to address precisely this limitation (Straus, Glasziou, Richardson, & Haynes, 2011).

Therefore, DLRs, which are independent of base rates and easy to use in practice with the aid of a nomogram, were calculated for YMRS and CDRS-R quintiles. An example, visually combining prior probability (i.e., base rate) with DLRs to determine posterior probability (i.e., PPV), is plotted on a nomogram in Figure 1. The use of nomograms as evidence-based assessment tools has been shown to improve clinicians' accuracy of diagnosis of pediatric BD and decrease variability among clinicians' diagnoses as well as decrease over-diagnosis of BD (Jenkins, Youngstrom, Feeny, Findling, & Youngstrom, 2011; Jenkins, Youngstrom, Washburn, & Youngstrom, 2011). Moreover, nomograms were rated by clinicians as easy to use, and a majority of clinicians said they would use them in their practice after receiving feedback about their improved accuracy (Jenkins, Youngstrom, Feeny, Findling, & Youngstrom, 2011).

The change in odds of a BD diagnosis was increased in the very high range of the YMRS for all comparisons. Given a YMRS score in the highest quintile when comparing BD to all other disorders, the probability of a BD diagnosis increased from the study base rate of 23% to just over 60%. Change in odds of a DD or mood diagnosis is slightly increased in the very high range of the CDRS-R for all comparisons.

DLRs can be applied across settings with different base rates with the use of a nomogram; however, it is important to note that the YMRS and CDRS-R, even with use of a nomogram, are not sufficient for making a diagnosis of BD or DD. YMRS and CDRS-R scores and use of a nomogram can inform the diagnostic process but cannot replace more thorough assessment of DSM-IV (or DSM-5) symptoms, family history, life events, and other contextual factors.

Differences on average YMRS and CDRS-R total scores do not discount the substantial nosological overlap among disorders, but speak to the ability of the YMRS and CDRS-R to differentiate groups on factors beyond nosological similarity and highlight the possibility that there are qualitative differences on overlapping symptoms between mood disorders and other disorders. Mick, Spencer, Wozniak, & Biederman (2005) posited irritability in BD is qualitatively different and more severe than other forms of irritability. Although not examined in detail here, perhaps other items common across disorders are qualitatively different. For example, “talks a lot” could be endorsed for either ADHD or mania, but chronic talkativeness versus episodic periods of pressured speech might be qualitatively different in ways that more nuanced assessments could distinguish.

Overall, the YMRS and CDRS-R, when administered in an unfiltered manner, have moderate to good discriminative validity and are able to increase or decrease the odds of a BD or DD diagnosis given scores within the very high or very low ranges of the scales. Nosological overlap among disorders did not eliminate the discriminative validity of the YMRS or CDRS-R. These findings are particularly encouraging considering the importance of accurate and effective outcome measures and the widespread use of the YMRS and CDRS-R in clinical research on pediatric BD.

All participants were treatment-seeking families at outpatient clinics associated with four universities conducting the LAMS study, and participants were screened with the intent of increasing the odds of BD in this sample. Therefore, the current study’s sample is enriched for BD, which could inflate diagnostic performance statistics, although statistics theoretically independent of base rate (e.g., DLRs) should remain generalizable (Zhou et al., 2002).

Raters were specifically instructed and trained to rate the YMRS and CDRS-R in an unfiltered manner. However, the same rater administered the YMRS, CDRS-R, and KSADS-PL-W. Raters administered the YMRS and CDRS-R before the KSADS-PL-W, and therefore were blind to diagnosis during their administration but not by the end of the interview. Future prospective studies should examine discriminative validity of unfiltered YMRS and CDRS-R administration with raters completely blind to diagnoses (Bossuyt et al., 2003). In addition, replication of this study in an adolescent population would allow for comparison of scores between the adolescent manifestation of mood, anxiety, behavior, and autism symptoms as well as other conditions more common to this developmental phase (e.g., substance abuse).

The ability of the YMRS and CDRS-R to differentiate BD and DD from other disorders does not negate the substantial nosological overlap between mood disorders and other disorders. Therefore, more work is needed to understand and flesh out the differences and similarities among diagnostic groups on the YMRS and CDRS-R, potentially with item analysis of unfiltered YMRS and CDRS-R administrations.

It is important to note that these results do not support the use of unfiltered administration of the YMRS and CDRS-R as diagnostic tools. These measures were not intended to be diagnostic tools and do not cover all diagnostic criteria, including all DSM-IV symptoms,

duration, and alternative explanations or causes for symptoms. Diagnoses remain guided by DSM-IV (or DSM-5) and should be informed by additional information not captured in the YMRS and CDRS-R, such as family history, medical history, traumatic life events, as well as onset and course of symptoms (e.g, Youngstrom, Findling, Youngstrom, & Calabrese, 2005; AACAP, 2007; Fields & Fristad, 2009; Youngstrom, Freeman, & Jenkins, 2009).

Filtered YMRS and CDRS-R administration (Youngstrom et al., 2005; Fraizer et al., 2007) achieved better discriminative validity than the current results for unfiltered administration. Therefore, filtered administration of these measures is still recommended for highest quality performance. However, current data support the use of unfiltered administration of the YMRS and CDRS-R in situations when it is unfeasible, due to lack or time or training, to conduct a KSADS or other diagnostic interview before assessing mood symptoms. The use of unfiltered YMRS and CDRS-R administration may also be particularly useful in clinical or research settings where symptom presentation is of foremost interest.

These instruments provide a useful means to monitor symptom severity over time and to measure treatment response. In addition, when used in a new evaluation, as described above, low scores indicate that more extensive evaluation of mood symptoms is not likely needed whereas high scores would direct the clinician to more thoroughly determine presence/absence of depressive or bipolar spectrum disorders. This is similar to good clinical practice using screening instruments for anxiety, disruptive behavior or autism – low scores decrease the need for further exploration of the disorder in question whereas high scores invite further investigation (Youngstrom, Choukas-Bradley, Calhoun, & Jensen-Doss, 2014).

Although results need to be replicated, initial results provide support for good discriminative validity of unfiltered administration of the YMRS and CDRS-R, which is encouraging given their use in the rapidly growing area of research on pediatric BD.

Acknowledgments

We would like to acknowledge the contributions of participant families as well as the Longitudinal Assessment of Manic Symptoms (LAMS) Group: L. Eugene Arnold, David Axelson, Christine Demeter, Thomas Frazier, Sarah M. Horwitz, Robert Kowatch, and H. Gerry Taylor. This study was supported by the National Institute of Mental Health (NIMH) (R01-MH073967; R01-MH073801; R01-MH073953; R01-MH073816). We thank NIMH for their support, but acknowledge that the findings and conclusions presented in this paper are those of the authors alone, and do not necessarily reflect the opinions of NIMH.

References

Allison, PD. Missing data. Sage; Thousand Oaks: 2002.

American Academy of Child and Adolescent Psychiatry (AACAP). Practice parameter for children and adolescents with bipolar disorder. Journal of the American Academy of Child and Adolescent Psychiatry. 2007; 46:107–125. http://www.jaacap.com/. [PubMed: 17195735]

Axelson D, Birmaher B, Strober M, Gill MK, Valeri S, Chiappetta L, Ryan N, Leonard H, Hunt J, Iyengar S, Bridge J, Keller M. Phenomenology of children and adolescents with bipolar spectrum disorders. Archives of General Psychiatry. 2006; 63:1139–1148.10.1001/archpsyc.63.10.1139 [PubMed: 17015816]

Blader JC, Carlson GA. Increased rates of bipolar disorder diagnoses among U.S. child, adolescent, and adult inpatients, 1996 – 2004. Biological Psychiatry. 2007; 62:107–114.10.1016/j.biopsych.2006.11.006 [PubMed: 17306773]

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, …de Vet HCW. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. British Medical Journal. 2003; 326:41–44.10.1136/bmj.326.7379.41 [PubMed: 12511463]

Carlson GA, Meyer SE. Phenomenology and diagnosis of bipolar disorder in children, adolescents, and adults: Complexities and developmental issues. Development and Psychopathology. 2006; 18:939–969.10.1017/S095457906060470 [PubMed: 17064424]

Carlson. Consensus report: Methodological issues and controversies in clinical trials with child and adolescent patients with bipolar disorder: Report of a consensus conference. Journal of Child and Adolescent Psychopharmacology. 2003; 13(1):13–27.10.1089/104454603321666162 [PubMed: 12804123]

Chambers WJ, Puig-Antich J, Hirsch M, Paez P, Ambrosini PJ, Tabrizi MA, Davies M. The assessment of affective disorders in children and adolescents by semistructured interview. Archives of General Psychiatry. 1985; 42:696–702. [PubMed: 4015311]

Costello JE, Erkanli A, Angold A. Is there an epidemic of child or adolescent depression? Journal of Child Psychology and Psychiatry. 2006; 47(12):1263–1271. [PubMed: 17176381]

Cummings, CM.; Caporino, NE.; Kendall, PC. Psychological Bulletin. 2013. Comorbidity of Anxiety and Depression in Children and Adolescents: 20 Years After.

Danner S, Fristad MA, Arnold LE, Youngstrom EA, Birmaher B, Horwitz SM, Demeter C, Findling RL, Kowatch RA. Early-onset bipolar spectrum disorders: Diagnostic issues. Clinical Child and Family Psychology Review. 2009; 12:27–293.

Fields BW, Fristad MA. Assessment of childhood bipolar disorder. Clinical Psychology: Science and Practice. 2009; 16(2):166–181.10.1111/j1468-2850.2009.01156.x

Findling RL, Gracious BL, McNamara NK, Youngstrom EA, Demeter CA, Branicky LA, Calabrese JR. Rapid, continuous cycling and psychiatric comorbidity in pediatric bipolar I disorder. Bipolar Disorders. 2001; 3:202–210. [PubMed: 11552959]

Findling RL, Youngstrom EA, Danielson CK, DelPorto-Bedoya D, Papish-David R, Townsend L, Calabrese JR. Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. Bipolar Disorders. 2002; 4(1):34–42.10.1034/j.1399-5618.2002.40102.x [PubMed: 12047493]

Findling RL, Youngstrom EA, Fristad MA, Birmaher B, Kowatch RA, Arnold E, Frazier TW, Axelson D, Ryan N, Demeter CA, Gill MK, Fields B, Depew J, Kennedy AM, Marsh L, Rowles BM, Horwitz SM. Characteristics of children with elevated symptoms of mania: The longitudinal assessment of manic symptoms (LAMS) study. Journal of Clinical Psychiatry. 2010; 71(12):1664–1672.10.4088/JCP.09m05859yel [PubMed: 21034685]

Frazier TW, Demeter CA, Youngstrom EA, Calabrese JR, Stansbrey RJ, McNamara NK, Findling RL. Evaluation and comparison of psychometric instruments for pediatric bipolar spectrum disorders in four age groups. Journal of Child and Adolescent Psychopharmacology. 2007; 17(6):853–866.10.1089/cap.2007.0057 [PubMed: 18315456]

Fristad MA, Algorta AP. Future directions for research on youth with bipolar spectrum disorders. Journal of Clinical Child & Adolescent Psychology. 2013; 42(5):734–747.10.1080/15374416.2013.817312 [PubMed: 23915232]

Fristad MA, Weller EB, Weller RA. The mania rating scale: can it be used in children? A preliminary report. Journal of the American Academy of Child and Adolescent Psychiatry. 1992; 31(2):252–256.10.1097/00004583-199203000-00011 [PubMed: 1564026]

Fristad MA, Weller EB, Weller RA. The mania rating scale (MRS): Further reliability and validity studies with children. Annals of Clinical Psychiatry. 1995; 7(3):127–132.10.3109/10401239509149039 [PubMed: 8646272]

Geller, B.; Craney, JL.; Bolhofner, K.; DelBello, MP.; Axelson, D.; Luby, J. Phenomenology and longitudinal course of children with a prepubertal and early adolescent bipolar disorder phenotype. In: Geller, B.; DelBello, MP., editors. Bipolar disorder in childhood and early adolescence. New York: Guilford Press; 2003. p. 25-50.

Geller B, Warner K, Williams M, Zimerman B. Prepubertal and young adolescent bipolarity versus ADHD: assessment and validity using the WASH-U-KSADS, CBCL and TRF. Journal of Affective Disorders. 1998; 51:93–100.10.1016/S0165-0327(98)00176-1 [PubMed: 10743842]

Geller B, Zimerman B, Williams M, Bolhofner K, Craney JL, DelBello MP, Soutullo C. Reliability of the Washington university in St. Louis kiddie schedule for affective disorders and schizophrenia (WASH-U-KSADS) mania and rapid cycling sections. Journal of the American Academy of Child and Adolescent Psychiatry. 2001; 40(4):450–455.10.1097/00004583 [PubMed: 11314571]

Goldstein, BI.; Liu, SM.; Schaffer, A.; Sala, R.; Blanco, C. Bipolar Disorders. 2013. Obesity and the three-year longitudinal course of bipolar disorder.

Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983; 148:839–843. [PubMed: 6878708]

Horwitz SM. Longitudinal assessment of manic symptoms (LAMS) study: background, design, and initial screening results. Journal of Clinical Psychiatry. 2010; 71(11):1511–1517.10.4088/JCP.09m05835yel [PubMed: 21034684]

Jenkins MM, Youngstrom EA, Feeny NC, Findling RL, Youngstrom JK. Generalizability of evidence-based assessment recommendations for pediatric bipolar disorder. Psychological Assessment. 2011 Oct 17. Advance online publication. 10.1037/a0025775

Jenkins MM, Youngstrom EA, Washburn JJ, Youngstrom JK. Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. Professional Psychology: Research and Practice. 2011; 42(2):121–129.10.1037/a0022506

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. Schedule for affective disoders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36(7):980–988. http://www.jaacap.com/. [PubMed: 9204677]

Kowatch R, Youngstrom EA, Danielyan A, Findling RL. Review and meta-analysis of the phenomenology and clinical characteristics of mania in children and adolescents. Bipolar Disorders. 2005; 7:483–496.10.1111/j1399-5618.2005.00261.x [PubMed: 16403174]

Lewinsohn PM, Klein DN, Seeley JR. Bipolar disorder in a community sample of older adolescents: Prevalence, phenomenology, comorbidity, and course. Journal of the American Academy of Child and Adolescent Psychiatry. 1995; 34:454–463.10.1097/00004583-199504000 [PubMed: 7751259]

Lewinsohn, PM.; Seeley, JR.; Klein, DN. Epidemiology and suicidal behavior. In: Geller, B.; DelBello, MP., editors. Bipolar disorder in childhood and early adolescence. New York: Guilford Press; 2003. p. 25-50.

Lonigan CJ, Phillips B, Hooe E. Relations of positive and negative affectivity to anxiety and depression in children: Evidence from a latent variable longitudinal study. Journal of Consulting and Clinical Psychology. 2003; 71:465–481. [PubMed: 12795571]

Merikangas KR, Pato M. Recent developments in the epidemiology of bipolar disorder in adults and children: Magnitude, correlates, and future directions. Clinical Psychology: Science and Practice. 2009; 16(2):121–133.10.1111/j.1468-2850.2009.01152.x

Metz CE. Investigative Radiology. 1986; 21:720–733.10.1097/00004424-198609000-00009 [PubMed: 3095258]

Mick E, Spencer T, Wozniak J, Biederman J. Heterogeneity of irritability in attention-deficit/hyperactivity disorder with and without mood disorders. Biological Psychiatry. 2005; 58:576–582.10.1016/j.biopsych.2005.05.037 [PubMed: 16084859]

Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. Annual Review of Psychology. 1998; 49:377–412.10.1146/annurev.psych.49.1.377

Moreno C, Laje G, Blanco C, Jiang H, Schmidt AB, Olfson M. National trends in the outpatient diagnosis and treatment of bipolar disorder in youth. Archives of General Psychiatry. 2007; 64(9):1032–1039.10.1001/archpsyc.64.9.1032 [PubMed: 17768268]

Poznanski EO, Freeman LN, Mokros HB. Children's depression rating scale – revised. Psychopathology Bulletin. 1985; 12(4):979–989.

Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, Gibbons R. Preliminary studies of the reliability and validity of the children's depression rating scale. Journal of the American Academy of Child and Adolescent Psychiatry. 1984; 23(2):191–197.10.1097/00004583-198403000-00011

Serrano E, Ezpeleta L, Alda JA, Matali JL, San L. Psychometric properties of the young mania rating scale for the identification of mania symptoms in Spanish children and adolescents with attention

deficit-hyperactivity disorder. Psychopathology. 2011; 44:125–132.10.1159/000320893 [PubMed: 21228617]

Siegel, S.; Castellan, NJJ. Nonparametric statistics for the behavioral sciences. 2. Boston: McGraw-Hill; 1988.

Straus, SE.; Glasziou, P.; Richardson, WS.; Haynes, RB. Evidence-based medicine: How to practice and teach EBM. 4. New York, NY: Churchill Livingstone; 2011.

Soutullo CA, Chang KD, Diez-Suárez A, Figueroa-Quintana A, Escamilla-Canales I, Rapado-Castro M, Ortuña F. Bipolar disorder in children and adolescents: International perspective on epidemiology and phenomenology. Bipolar Disorders. 2005; 7:497–506.10.1111/j1399-5618.2005.00262.x [PubMed: 16403175]

Swets K. Measuring the accuracy of diagnostic systems. Science. 1988; 240:1285–1293.10.1126/science.3287615 [PubMed: 3287615]

Van Meter AR, Moreira ALR, Youngstrom EA. Meta-analysis of epidemiologic studies of pediatric bipolar disorder. Journal of Clinical Psychiatry. 2011; 72(9):1250–1256.10.4088/JCP.10m06290 [PubMed: 21672501]

World Health Organization (WHO). World report on disability. Geneva, Switzerland: WHO; 2011.

Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. British Journal of Psychiatry. 1978; 133:429–435.10.1192/bjp.133.5.429 [PubMed: 728692]

Youngstrom, EA.; Choukas-Bradley, S.; Calhoun, CD.; Jensen-Doss, A. Cognitive and Behavioral Practice. 2014. Clinical guide to the Evidence-Based Assessment approach to diagnosis and treatment.

Youngstrom EA, Danielson CK, Findling RL, Gracious BL, Calabrese JR. Factor structure of the young mania rating scale for use with youths ages 5 to 17 years. Journal of Clinical Child and Adolescent Psychology. 2002; 31(4):567–572.10.1207/S15374424JCCP3104_15 [PubMed: 12402575]

Youngstrom EA, Duax J. Evidence-based assessment of pediatric bipolar disorder, part I: Base rate and family history. Journal of the American Academy of Child and Adolescent Psychiatry. 2005; 44(7):712–717. http://www.jaacap.com/. [PubMed: 15968241]

Youngstrom EA, Findling RL, Calabrese JR, Gracious BL, Demeter CA, Bedoya DD, Price M. Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. Journal of the American Academy of Child and Adolescent Psychiatry. 2004; 43(7):847–858.10.1097/01.chi.0000125091.35109.1e [PubMed: 15213586]

Youngstrom EA, Findling RL, Danielson CK, Calabrese JR. Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. Psychological Assessment. 2001; 13(2):267–276.10.1037//1040-3590.13.2.267 [PubMed: 11433802]

Youngstrom EA, Findling RL, Youngstrom JK, Calabrese JR. Toward an evidenced based assessment of pediatric bipolar disorder. Journal of Clinical Child and Adolescent Psychology. 2005; 34(3):433–448. http://www.jaacap.com/. [PubMed: 16026213]

Youngstrom EA, Frazier TW, Demeter CA, Calabrese JR, Findling RL. Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. The Journal of Clinical Psychiatry. 2008; 69(5):831–839.10.4088/JCP.v69n0517 [PubMed: 18452343]

Youngstrom EA, Freeman AJ, Jenkins MM. The assessment of bipolar disorder in children and adolescents. Child and Adolescent Psychiatric Clinics of North America. 2009; 18(2):353–ix.10.1016/j/chc.2008.12.002 [PubMed: 19264268]

Youngstrom EA, Meyers O, Demeter C, Youngstrom J, Morello L, Piiparinen R, Feeny N, Calabrese JR, Findling RL. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. Bipolar Disorders. 2005; 7:507–517.10.1111/j.1399-5618.2005.00269.x [PubMed: 16403176]

Youngstrom EA, Meyers O, Youngstrom JK, Calabrese JR, Findling RL. Comparing the effects of sampling designs on the diagnostic accuracy of eight promising screening algorithms for pediatric bipolar disorder. Biological Psychology. 2006; 60:1013–1019.10.1016/j.biopsych.2006.06.023

Youngstrom EA, Meyers O, Youngstrom JK, Calabrese JR, Findling RL. Diagnostic and measurement issues in the assessment of pediatric bipolar disorder: Implications for understanding mood

disorder across the life cycle. Development and Psychophathology. 2006a; 18:989–1021.10.1017/S0954579406060494

Zhou, X-H.; Obuchowski, NA.; McClish, DK. Statistical methods in diagnostic medicine. New York: Wiley; 2002.

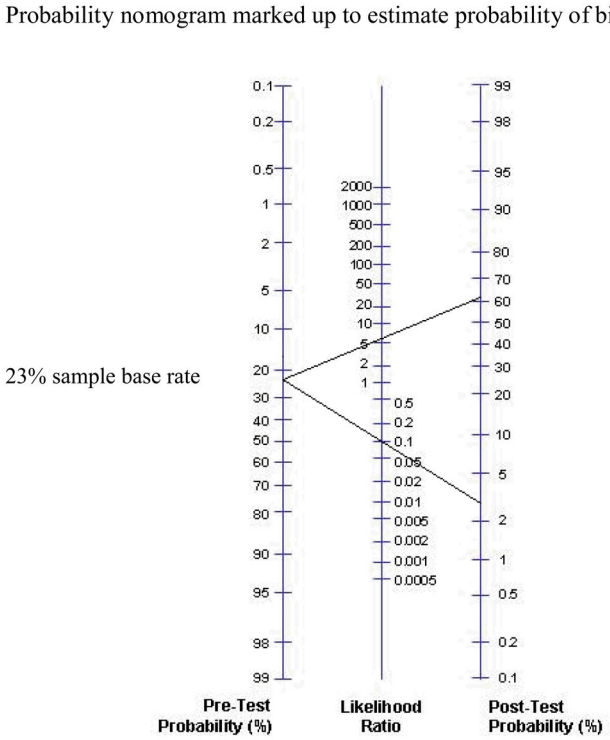

Figure 1.

Probability nomogram marked up to estimate probability of bipolar diagnosis

An example of how to use a nomogram for the YMRS in the current sample. The first column is the study base-rate for BD, the second column is the calculated DLR for quintiles of the YMRS, and the third column is PPV, which can be interpreted as the probability of having BD if a test result falls within the corresponding quintile. DLRs corresponding to the highest and lowest YMRS quintiles are plotted for the BD vs. No BD comparison. The first column base-rate changes for different settings and local base-rates should be used.

Table 1

Mania Symptom Overlap with Other Common Disorders

Symptoms of Mania	ADHD	ODD	CD	ASD	ANX
Elevated/expansive mood	May act silly, potentially due to lack of inhibition			Can present in a silly, giggly manner	
Irritable mood	Can be irritable	Diagnostic criterion: temper outbursts, easily annoyed, and angry/resentful	Can have extreme irritability	Can have extreme irritability	Diagnostic criterion for some anxiety disorders and present in many anxiety disorders
Inflated self-esteem or grandiosity	May exaggerate their abilities relative to reality	Many times they do not think about the consequences of their actions and disrespect authority, which can present as inflated self-esteem	Disrespecting authority and violating others' rights, which can present as inflated self-esteem	May lack fear in response to dangers that could appear grandiose	
Decreased need for sleep	Some sleep less than their same-aged peers; also may be a side effect of stimulant medication or a result of dose taken late in the day			Many wake at night or very early in the AM without showing expected effects of sleep loss	Many have sleep disruption
More talkative/pressured speech	Diagnostic criterion for: talks excessively			Verbal rituals may present like pressured speech	Speaking quickly when nervous
Flight of ideas/racing thoughts	Can have deficits in inhibition and functional impairment in working memory and internalization of speech, which could present like flight of ideas or racing thoughts			Often have expressive language difficulties making it difficult to follow their train of thought or are guided by preoccupations	May endorse having thoughts they cannot stop
Distractibility	Diagnostic criterion for ADHD			Can be interested in internal stimuli or unusual sensory perceptions and may appear to be distracted easily	May have a hard time concentrating due to preoccupations
Increase in goal directed activity/ psychomotor agitation	Often fidget, pace, and move around a lot				May have trouble sitting still, may fidget and feel restless
Excessive involvement in pleasurable activities with high potential for negative consequences	May present with this due to impulsivity and hyperactivity	Often act regardless of consequences	May present with this due to a pattern of dangerous, negative behaviors	May become absorbed in restricted interests that could have negative consequences	

Table 2

Depression Symptom Overlap with Other Common Disorders

Symptoms of Depression	ADHD	ODD	CD	ANX
Depressed mood				
Irritable mood	Can be irritable	Diagnostic criterion: temper outbursts, easily annoyed, and angry/ resentful	Can have extreme irritability	Diagnostic criterion for some and present in many anxiety disorders
Diminished pleasure or interest in activities	Do not enjoy usual activities at a usual level and can become easily bored			Sometimes avoid activities, which may look like diminished interest
Weight loss or gain/appetite decrease of increase	Stimulant medication for ADHD can cause appetite changes			May have physiological symptoms (e.g., nausea) that could lead to changes in appetite
Insomnia/hypersomnia	Some sleep less than their same-aged peers			Diagnostic criterion for GAD and PTSD and others can have trouble with sleep, particularly initial insomnia
Psychomotor agitation/retardation	Often fidget, pace, and move a lot			May have trouble sitting still, may fidget and feel restless
Fatigue/loss of energy				Often have initial insomnia and are tired the next day
Feelings of worthlessness/excessive guilt				May worry about imperfections and be self-conscious or may have excessive feelings of guilt
Difficulty concentrating	Diagnostic criterion			May report difficulty concentrating
Recurrent thoughts of death/suicidal ideation/ suicide attempt				May have morbid obsessions, particularly in OCD

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2016 November 01.

Table 3

Demographics for Participants with Bipolar Disorder (BD) and Comparison Groups

	BD group	Comparison groups					Total
Characteristic	**BD (*n*=162)**	**No BD (*n*=545)**	**ODD/CD (*n*=293)**	**ASD (*n*=32)**	**ADHD (*n*=157)**	**ANX (*n*=49)**	**(*N*=707)**
Age, mean (*SD*)	9.8 (2.1)	9.3 (1.9)**	9.1 (1.9)***	9.2 (1.7)	9.3 (1.9)*	9.7 (1.8)	9.4 (1.9)
Male, *n* (%)	92 (56.8)	386 (70.8)**	212 (72.4)**	30 (93.8)***	117 (74.5)**	21 (42.9)	478 (67.6)
Race, white, *n* (%)	88 (54.3)	367 (63.7)**	202 (68.9)**	21 (65.6)	106 (67.5)*	36(73.5)	455 (64.4)
Ethnicity-Hispanic, *n* (%)	8 (4.9)	23 (4.2)	12 (4.1)	0 (0.0)	10 (6.4)	1(2)	31 (4.4)
Medicaid, *n*, (%)	77 (47.5)	293 (53.8)	168 (57.3)	8 (25.0)*	88 (56.1)	17(34.7)	370 (52.3)
CGAS, mean (*SD*)	51.2 (9.4)	55.5 (10.4)***	53.2 (9.4)*	47.2 (10.2)*	60.2 (8.6)***	55.9 (10.9)**	54.6 (10.3)
YMRS, mean *(SD)*	25.6 (8.9)	14.2 (7.5)***	15.9 (7.3)***	15.0 (8.9)***	12.3 (6.7)***	10.9(1.5)***	16.8 (9.2)

Note: CGAS = Children's Global Assessment Scale; BD = Bipolar spectrum disorders; ODD = oppositional/defiant disorder; CD = conduct disorder; ASD = autistic spectrum disorder; ADHD = attention deficit hyperactivity disorder; ANX = Anxiety Disorders.

* $p < .05$;

** $p < .01$;

*** $p < .001$; compared to the BD group.

Table 4

Demographics for Participants with Depressive Disorders (DD) and Comparison Groups

BD included in comparison groups

	DD group	Comparison groups			Total
Characteristic	**DD (*n*=124)**	**No DD (*n*=583)**	**ODD/CD (*n*=292)**	**ANX (*n*=78)**	**(*N*=707)**
Age, mean (*SD*)	10.0 (1.8)	9.3 (1.9)***	9.1 (1.9)***	9.2 (1.8)**	9.4 (1.9)
Male, *n* (%)	80 (64.5)	398 (68.3)	204 (69.9)	49 (62.8)	478 (67.6)
Race, white, *n* (%)	61 (49.2)	394 (67.6)***	200 (68.5)***	53 (67.9)*	455 (64.4)
Ethnicity-Hispanic, *n* (%)	4 (3.2)	27 (4.6)	12 (4.1)	4 (5.1)	31 (4.4)
Medicaid, *n*, (%)	64 (51.6)	306 (52.5)	164 (56.2)	26 (33.3)*	370 (52.3)
CGAS, mean (*SD*)	50.8 (9.0)	55.4 (10.4)***	52.9 (9.9)*	55.5 (8.6)***	54.6 (10.3)
CDRS-R, mean *(SD)*	43.7 (10.4)	32.9 (9.8)***	34.0 (9.6)***	34.8 (9.8)***	34.8 (10.7)

BD excluded from comparison groups

	DD group	Comparison groups			Total
Characteristic	**DD (*n*=124)**	**No DD (*n*=421)**	**ODD/CD (*n*=224)**	**ANX (*n*=54)**	**(*N*=707)**
Age, mean (*SD*)	10.0 (1.8)	9.1 (1.8)***	8.9 (1.8)***	9.1 (1.8)**	9.4 (1.9)
Male, *n* (%)	80 (64.5)	306 (72.7)	163 (72.8)	37 (68.5)	478 (67.6)
Race, white, *n* (%)	61 (49.2)	306 (72.7)***	168 (75.0)***	38 (70.4)*	455 (64.4)
Ethnicity-Hispanic, *n* (%)	4 (3.2)	19 (4.5)	9 (4.0)	4 (7.4)	31 (4.4)
Medicaid, *n*, (%)	64 (51.6)	229 (54.4)	129 (57.6)	17 (31.5)*	370 (52.3)
CGAS, mean (*SD*)	50.8 (9.0)	56.9 (10.3)***	54.5 (9.5)***	56.7 (9.0)***	54.6 (10.3)
CDRS-R, mean *(SD)*	43.7 (10.4)	30.7 (8.2)***	31.6 (7.9)***	33.2 (9.1)***	34.8 (10.7)

BD and DD combined into a mood group

	Mood group	Comparison groups			Total
Characteristic	**Mood (*n*=286)**	**No Mood (*n*=421)**	**ODD/CD (*n*=224)**	**ANX (*n*=54)**	**(*N*=707)**
Age, mean *(SD)*	9.9 (2.0)	9.1 (1.8)***	8.9 (1.8)***	9.1 (1.8)**	9.4 (1.9)
Male, *n (%)*	172 (60.1)	306 (72.7)**	163 (72.8)**	37 (68.5)	478 (67.6)
Race, white, *n (%)*	149 (52.1)	306 (72.7)***	168 (75.0)***	38 (70.4)*	455 (64.4)
Ethnicity-Hispanic, *n (%)*	12 (4.2)	19 (4.5)	9 (4.0)	4 (7.4)	31 (4.4)
Medicaid, *n, (%)*	141 (49.3)	229 (54.4)	129 (57.6)	17 (31.5)*	370 (52.3)
CGAS, mean *(SD)*	51.0 (9.2)	56.9 (10.3)***	54.5 (9.5)***	56.7 (9.0)***	54.6 (10.3)
CDRS-R, mean *(SD)*	41.0 (11.0)	30.7 (8.2)***	31.6 (7.9)***	33.2 (9.1)***	34.8 (10.7)

Note: CGAS = Children's Global Assessment Scale; BD = bipolar disorder; DD = unipolar depressive spectrum disorders; ODD = oppositional/defiant disorder; CD = conduct disorder; ANX = anxiety disorders.

* $p < .05$;

** $p < .01$;

*** $p < .001$; compared to the DD group.

Table 5

Comorbid Diagnoses for Participants with Bipolar Disorder (BD) and Comparison Groups

Diagnosis, n (%)	BD group	Comparison groups				
	BD (n=162)	No BD (n=545)	ODD/CD (n=293)	ASD (n=32)	ADHD (n=157)	ANX (n=49)
DD	NA[a]	124 (22.8)	70 (23.9)	2 (6.3)	32 (20.4)	21(42.8)
ADHD	117 (72.2)	421 (77.2)	247 (84.3)	17 (53.1)	*157 (100)*	NA[b]
ASD	5 (3.1)	40 (7.3)	8 (2.7)	*32 (100)*	NA[b]	NA[b]
ODD/CD	68 (42.0)	293 (53.8)	*293 (100)*	NA[b]	NA[b]	NA[b]
ANX	65 (40.1)	133 (24.4)	112 (38.2)	13 (40.6)	62 (39.5)	*49 (100)*
Other	50 (30.9)	182 (33.4)	99 (33.8)	12 (37.5)	35 (22.3)	25(51)

Note: BD = Bipolar spectrum disorders; ODD = oppositional/defiant disorder; CD = conduct disorder; ASD = autistic spectrum disorder; ADHD = attention deficit hyperactivity disorder; ANX = anxiety disorders.

[a] Depressive disorders cannot be comorbid with bipolar disorders in the DSM nosology.

[b] These comorbid diagnoses were not allowed within the hierarchical division of comparison groups for analyses.

Table 6

Comorbid Diagnoses for Participants with Depressive Disorders (DD) and Comparison Groups

BD included in comparison groups

Diagnosis, n (%)	DD group	Comparison group		
	DD (n=124)	No DD (n=583)	ODD/CD (n=292)	ANX (n=78)
BD	NA[a]	162 (27.8)	68 (23.3)	24 (30.8)
ADHD	95 (76.6)	443 (76.0)	247 (84.6)	46 (59.0)
ASD	2 (1.6)	43 (7.4)	11 (3.8)	9 (11.5)
ODD/CD	69 (55.6)	292 (50.1)	*292 (100)*	NA[b]
ANX	88 (71.0)	210 (36.0)	109 (37.3)	*78 (100)*
Other	37 (29.8)	195 (33.4)	111 (38.0)	24 (30.8)

BD excluded from comparison groups

Diagnosis, n (%)	DD group	Comparison group		
	DD (n=124)	No DD (n=421)	ODD/CD (n=224)	ANX (n=54)
BD	NA[a]	NA[c]	NA[c]	NA[c]
ADHD	95 (76.6)	326 (77.4)	197 (87.9)	35 (64.8)
ASD	2 (1.6)	38 (9.0)	8 (3.6)	9 (16.7)
ODD/CD	69 (55.6)	224 (53.2)	*224 (100)*	NA[b]
ANX	88 (71.0)	145 (34.4)	73 (32.6)	*54 (100)*
Other	37 (29.8)	145 (34.4)	84 (37.5)	17 (31.5)

DD and BD combined into a mood group

Diagnosis, n (%)	Mood group	Comparison groups		
	Mood (n=286)	No Mood (n=421)	ODD/CD (n=224)	ANX (n=54)
BD	*162 (56.6)*[d]	NA[b]	NA[b]	NA[b]
DD	*124 (43.3)*[d]	NA[b]	NA[b]	NA[b]
ADHD	212 (74.1)	326 (77.4)	197 (87.9)	35 (64.8)

DD and BD combined into a mood group

Diagnosis, n (%)	Mood group	Comparison groups		
	Mood (n=286)	No Mood (n=421)	ODD/CD (n=224)	ANX (n=54)
ASD	7 (2.4)	38 (9.0)	8 (3.6)	9 (16.7)
ODD/CD	137 (47.9)	224 (53.2)	*224 (100)*	NA[b]
ANX	153 (53.5)	145 (34.4)	73 (32.6)	*54 (100)*
Other	87 (30.4)	145 (34.4)	84 (37.5)	17 (31.5)

Note: BD = Bipolar spectrum disorders; DD = unipolar depressive spectrum disorders; ODD = oppositional/defiant disorder; CD = conduct disorder; ANX = anxiety disorders.

[a] Bipolar disorders cannot be comorbid with depressive disorders in the DSM nosology.

[b] These comorbid diagnoses were not allowed within the hierarchical division of comparison groups for analyses.

[c] BD was excluded from all comparison groups.

[d] BD and DD comprised the Mood group, and therefore, are not comorbid diagnoses.

Table 7

Discriminative Validity Analyses for YMRS and CDRS-R

YMRS							
Comparison	**AUC**	**95% CI**	**Optimal Cut Pt.**	**Sens**	**Spec**	**PPV[a]**	**NPV[a]**
BD vs. No BD	.84	.80–.87	22.5	.68	.86	.71 (.23)	.90 (.23)
BD vs. ODD/CD	.80	.76–.84	22.5	.68	.82	.67 (.36)	.82 (.36)
BD vs. ASD	.80	.72–.88	22.5	.68	.81	.95 (.84)	.33 (.84)
BD vs. ADHD	.88	.84–.92	19.5	.75	.87	.86 (.51)	.77 (.51)
BD vs. ANX	.90	.86–.95	17.5	.82	.86	.95 (.77)	.55 (.77)
CDRS-R							
	BD included in comparison groups						
Comparison	**AUC**	**95% CI**	**Optimal Cut Pt.**	**Sens**	**Spec**	**PPV[a]**	**NPV[a]**
DD vs. No DD	.78	.74–.82	38.5	.73	.74	.35 (.18)	.93 (.18)
DD vs. ODD/CD	.76	.72–.81	38.5	.73	.73	.53 (.30)	.86 (.30)
DD vs. ANX	.73	.66–.80	38.5	.73	.65	.77 (.61)	.60 (.61)
	BD excluded from comparison groups						
DD vs. No DD	.84	.80–.88	36.5	.76	.79	.52 (.23)	.92 (.23)
DD vs. ODD/CD	.83	.78–.87	38.5	.73	.82	.69 (.36)	.84 (.36)
DD vs. ANX	.78	.70–.85	34.5	.82	.63	.84 (.70)	.60 (.70)
	BD and DD combined into a mood group						
Mood vs. No mood	.77	.74–.81	36.5	.63	.79	.67 (.40)	.76 (.40)
Mood vs. ODD/CD	.75	.71–.79	36.5	.63	.79	.79 (.56)	.62 (.56)
Mood vs. ANX	.71	.63–.78	33.5	.72	.59	.90 (.84)	.29 (.84)

Note: AUC = area under the curve; CI = confidence interval; Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value. The “optimal cutpoint” maximized the combination of sensitivity and specificity in the sample. BD = Bipolar spectrum disorders; DD = unipolar depressive spectrum disorders; ODD = oppositional/defiant disorder; CD = conduct disorder; ASD = autistic spectrum disorder; ADHD = attention deficit hyperactivity disorder; ANX = anxiety disorders.

[a]Study base rate is presented for each comparison is presented in parentheses following PPV and NPV.

Table 8

Change in Odds of Diagnosis (Diagnostic Likelihood Ratios) for YMRS

	YMRS				
	Score range				
	Very Low	**Low**	**Neutral**	**High**	**Very High**
YMRS score:	**≤2**		**9–18**	**19–24**	**25+**
BD vs. No BD	.10		.41[a]	1.12	6.12
YMRS score:	**≤10**		**11–21**	**22–27**	**28+**
BD vs. ODD/CD	.19		.49[a]	1.76	5.74
YMRS score:	**≤9**		**10–21**	**22–28**	**29+**
BD vs. ADHD	.09		.50[a]	5.60	11.82
YMRS score	**≤7**		**8–16**	**17–23**	**24+**
BD vs. ANX	.09		.76	2.94	13.91

Note: BD = Bipolar spectrum disorders; DD = unipolar depressive spectrum disorders; ODD = oppositional/defiant disorder; CD = conduct disorder; ADHD = attention deficit hyperactivity disorder; ANX = anxiety disorders.

[a]DLRs that were similar across two or more score ranges were collapsed.

Table 9

Change in Odds of Diagnosis (Diagnostic Likelihood Ratios) for CDRS-R

BD included in comparison groups					
	Score range				
	Very Low	Low	Neutral	High	Very High
CDRS-R score:	**≤25**		**26–35**	**36–44**	**45+**
DD vs. No DD	.13		.42[a]	1.76	3.12
CDRS-R score:	**≤27**		**28–38**	**39–46**	**47+**
DD vs. ODD/CD	.21		.47[a]	2.20	3.30
CDRS-R score:	**≤30**		**31–49**		**50+**
DD vs. ANX	.26		1.14[a]		3.56

BD excluded from comparison groups					
	Score range				
	Very Low	Low	Neutral	High	Very High
CDRS-R score:	**≤29**		**30–41**		**42+**
DD vs. No DD	.15[a]		.93[a]		5.08
CDRS-R score:	**≤26**		**27–37**	**38–44**	**45+**
DD vs. ODD/CD	.12		.42[a]	2.36[a]	6.07
CDRS-R score:	**≤30**		**31–50**		**51+**
DD vs. ANX	.23		1.15[a]		14.37

BD and DD combined to form mood group					
	Score range				
	Very Low	Low	Neutral	High	Very High
CDRS-R Scores	**≤25**	**26–30**	**31–35**	**36–44**	**45+**
Mood vs. No mood	.25	.41	.78	1.53	5.63
CDRS-R Scores	**≤27**	**28–32**	**33–38**	**39–46**	**47+**

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2016 November 01.

BD and DD combined to form mood group

	Score range				
	Very Low	**Low**	**Neutral**	**High**	**Very High**
Mood vs.ODD/CD	.39	.49	.70	1.97	7.56
CDRS-R Scores	**≤29**	**30–36**	**37–42**	**43–49**	**50+**
Mood vs. ANX	.44	.81	1.12	1.57	3.92

Note: BD = Bipolar spectrum disorders; DD = unipolar depressive spectrum disorders; ODD = oppositional/defiant disorder; CD = conduct disorder; ANX= anxiety disorders.

[a] DLRs that were similar across two or more score ranges were collapsed.