DIAGNOSTIC ACCURACY OF THE 7 UP 7 DOWN INVENTORY: DIFFERENTIATING UNIPOLAR AND BIPOLAR DEPRESSION IN AN OUTPATIENT SETTING

Tate Halverson

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in The Department of Psychology and Neuroscience (Clinical Psychology).

Chapel Hill
2017

Approved by:

David Penn

Eric Youngstrom

Mitch Prinstein

Jennifer Youngstrom

ABSTRACT

Tate Halverson: Diagnostic Accuracy of the 7 Up 7 Down Inventory: Differentiating Unipolar and Bipolar Depression in an Outpatient Setting
(Under the direction of Eric Youngstrom and David Penn)

This study examined the clinical utility of the 7 Up 7 Down Inventory (7U7D) in youths (ages 5-18; $N = 1737$) presenting to outpatient mental health clinics. Caregivers and youths completed the 7U7D and a semi-structured interview to determine psychiatric diagnoses. Caregiver and youth-reported 7U7D scores significantly identified youth mood and bipolar disorders (areas under the curve .56 - .81, ps $<.05$), with caregiver-report significantly outperforming youth-report. The 7U7D showed strong incremental validity after controlling for youth demographics and clinical characteristics. Cutoff scores were calculated to generate diagnostic likelihood ratios (DiLR) in a two-step fashion to utilize both hypomanic/manic (7U) and depression (7D) dimensions. 7D optimal cut scores yielded DiLRs between 1.55 and 3.26 for a mood disorder diagnosis. 7U Optimal cut scores yielded DiLRs between 1.00 and 2.11 for a BP diagnosis. The 7U7D demonstrates clinical utility for identifying youth mood disorders and BP from other psychiatric diagnoses.

ACKNOWLEDGEMENTS

I would like to thank my committee members Drs. David Penn, Eric Youngstrom, Mitch Prinstein, and Jennifer Youngstrom, the MECCA Lab (Stephanie Salcedo, Mian-Li Ong, Jacquee Genzlinger) and the Penn Lab (Julia Browne, Ben Buck, Emily Gagen, Kelsey Ludwig, Arundati Nagendra, Lana Nye, and Grace Lee Simmons), Dr. Robert Findling, Chris Wiesen at the Odum Institute, and the agencies supporting initial data collection (National Institute of Mental Health, R01 066647, PI Dr. Eric Youngstrom; Stanley Institute, PI Dr. Robert Findling).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

7D	7 Down Subscale
7U	7 Up Subscale
7U7D	7 Up 7 Down Inventory
ADHD	Attention Deficit/Hyperactivity Disorder
AUC	Area Under the Curve
BP	Bipolar Disorder
CD	Conduct Disorder
DiLR	Diagnostic Likelihood Ratio
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders– 4th Edition
GBI	General Behavior Inventory
K-SADS	The Kiddie Schedule for Affective Disorders and Schizophrenia
MDD	Major Depressive Disorder
ODD	Oppositional Defiant Disorder
PGBI-10M	Parent General Behavior Inventory – 10 – Item Mania Scale

INTRODUCTION

Bipolar disorder (BP) is a disorder characterized by fluctuations in mood (American Psychiatric Association, 2013). However, BP is not a heterogeneous diagnosis; the presence of just one manic episode in the absence of depression, depression with hypomania, and depression with subthreshold hypomania will all result in a diagnosis of BP. BP is comprised of different subtypes including Bipolar I (BPI), Bipolar II (BPII), cylothymia, and Other Specified Bipolar and Related Disorders (OS-BRD), formerly Not Otherwise Specified (BP-NOS). The different subtypes of BP are differentiated by intensity and duration of mood episodes and can be conceptualized as existing along a spectrum. Estimates for lifetime prevalence of BP range from 1% - 6.4% (Akiskal et al., 2000; Judd & Akiskal, 2003; Merikangas et al., 2007). Conservative prevalence estimates for BP are based on narrowly defined conceptualizations of BP (i.e., BPI and BPII) that rely on a minimum 4-day duration of hypomanic symptoms. However, several studies show brief episodes with mood symptoms lasting only 1-3 days are comparable in clinical significance to episodes lasting longer than 4 days and are associated with significant psychosocial consequences (Akiskal et al., 2000; Angst et al., 2002).

The average age of onset for BP is 18 - 22 years; however, studies suggest age of onset for BP may be much younger when using a less stringent diagnostic criterion for symptom duration (Merikangas et al., 2007). Axelson et al. (2006) found children and adolescents received a diagnosis of BP-NOS mainly because they did not meet the minimum symptom duration criteria for BPI or BPII diagnoses. However, these children and adolescents still exhibited symptom intensity requirements and a family history of BP, suggesting a continuum of BP

evident at a much younger age. Prospective follow-up studies support this idea of BP as a continuum along which individuals progress with age. Alloy et al. (2012) followed individuals with a diagnosis of BP-NOS and cyclothymia and found over 50% progressed to a BPI or BPII diagnosis whereas 17.4% of BPII diagnoses progressed to BPI. A meta-analysis examining BP in youths found the prevalence rate of pediatric BP spectrum diagnoses to be 1.8% (Van Meter, Moreira, & Youngstrom, 2011). Additionally, a five-year prospective study in youths found similar rates of conversion as adults with 45% of youths originally diagnosed with BP-NOS converting to BP-I or BP-II with an average time to conversion of 58 weeks (Axelson et al., 2011).

BP in youths and adults is associated with substantial impairment, economic costs, and suicide risk. Cross-sectional studies show large decreases in quality of life, especially in areas of social and emotional functioning as well as high comorbidity with alcohol and substance abuse (Akiskal et al., 2000; Simon, 2003). BP is also a pressing public health concern with substantially higher health care costs compared with other mental health diagnoses as well as significant increases in the need for public assistance (Judd & Akiskal, 2003). Pediatric BP in particular is associated with substantial impairment, including high rates of prior hospitalizations, depressive episodes, treatment with medication, comorbid anxiety disorders, and suicidal ideation (Axelson et al., 2006). Overshadowing the economic costs is the high rate of suicide prevalence in BP. Lifetime prevalence of suicide for BP is 8% with some research suggesting this estimate may be as high as 28.1% when including BPII, which has an especially elevated risk for suicide (Angst et al., 2002; Berk & Dodd, 2005).

Substantial impairment and economic costs highlight the need for accurate diagnosis of BP in order for individuals to receive appropriate treatment. Complicating accurate diagnosis of BP is the overlap in symptom presentation with major depressive disorder (MDD) or unipolar

depression and externalizing disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD) and Conduct Disorder (CD) (Biederman, 1998; Kim & Miklowitz, 2002; Youngstrom, Arnold, & Frazier, 2010) Misdiagnosis of BP causes significant delays between symptom onset and effective treatment, as well as missed early intervention windows and increased health care costs due to worsening of symptoms from inappropriate treatment approaches.

Misdiagnosis of Bipolar Disorder

Estimates for rates of misdiagnosis of BP in adults range from 3% - 66% with the majority of cases assigned an original diagnosis of a unipolar mood disorder (Berk et al., 2011; Keck, Kessler, & Ross, 2008; McCombs, Ahn, Tencer, & Shi, 2007; Smith et al., 2011). Smith et al. (2011) recruited patients diagnosed with MDD from a primary care setting for structured diagnostic interviews and confirmed the presence of BP in 1 out of 30 cases. The study was a two-part screening study and rates of misdiagnosis were estimated using both conservative (i.e., assuming participants not reached for a diagnostic interview would not meet criteria for BP) and more permissive methods (i.e., assuming participants not reached for a diagnostic interview would meet criteria for BP at the same rate as participants interviewed) yielding estimates for misdiagnosis of BP ranging from 3% to 21.6%. Individuals with BP that experience more depressive episodes than manic or hypomanic episodes also experience longer delays in accurate diagnosis with as many as 50% of BP cases initially classified as unipolar depression (Altamura et al., 2010; Wolkenstein, Bruchmuller, Schmid, & Meyer, 2011).

Adults, and possibly adolescents, with BP most commonly seek initial treatment for depressed symptoms and may not report manic or hypomanic symptoms (Bowden, 2001, 2005). A diagnosis of BP requires the presence of at least one manic or hypomanic episode, so without proper screening for lifetime manic or hypomanic symptoms, these individuals are often initially diagnosed with a unipolar mood disorder. Complicating the diagnostic picture is the course of

BP may begin with a depressed episode with manic or hypomanic symptoms experienced later in development (Hillegers et al., 2005). The presence of depressive symptoms *only* in the absence of clear manic symptoms is one major cause of misdiagnosis of BP, especially in early phases of the disorder (Arrasate et al., 2014). A further complication of diagnosing BP is that patients presenting for treatment while depressed underreport a history of hypomanic and manic symptoms since shifts in activity, energy, and sleep are seen as improvements compared with current depression symptoms (Angst et al., 2002).

Symptom Overlap

Symptom overlap between BP and unipolar depression, as well as difficulties with assessment of prior or current hypomanic and manic symptoms contribute to long delays between onset of affective symptoms and accurate diagnosis and appropriate treatment. The rate of delay in BP between symptom onset and appropriate treatment ranges from 6 to 12 years (Altamura et al., 2010; Berk et al., 2007). The rate of delay is estimated to be similar in children with delays of 5 to 12 years and correct diagnosis occurring in the first year of symptoms in only 4.8% of youth cases (Marchand, 2006). This delay in symptom presentation and appropriate treatment comes at a substantial cost to individuals experiencing impairment as well as previously discussed economic costs. Unrecognized BP is associated with significantly greater health care costs overall as well as increased costs per each year of unrecognized BP (Keck et al., 2008; McCombs et al., 2007). Prospective studies also find longer durations of improperly treated BP are associated with increased hospitalization rates as well as higher rates of suicide attempts and suicide (Altamura et al., 2010).

One source of misdiagnosis, especially in youths, is the presence of dysphoric mania and labile mood states commonly observed in youths with BP (Youngstrom, Boris Birmaher, & Findling, 2008). Youngstrom, Birmaher, and Findling (2008) describe the presence of two

phenomena – "chocolate milk" and "fudge ripple" – of clinical presentations seen in youths with BP, which increase the complexity of an accurate diagnosis. The chocolate milk presentation is seen when youths have simultaneous symptoms of both depression and mania, making it difficult to separate symptoms into clear episodes of depression *or* mania. The fudge ripple presentation is the presence of short, episodic, durations of mania or hypomania and depression that can occur several times per day. Rates of mood lability without interepisode recovery are observed in 50% of youth cases of BPI as well as a 75% comorbidity rate with at least one disruptive behavior disorder (Findling et al., 2001).

Comorbidity

High rates of comorbidity between BP and externalizing disorders and attention disorders are another source of misdiagnosis of BP in youths. Bowring and Kovacs (1992) present evidence for symptom overlap of mania with other disorders, such as ADHD and conduct disorder, as a main difficulty diagnosing BP in youths. Furthermore, there is also evidence for shared mechanisms (e.g., shared risk factors) leading to high rates of comorbidity with ADHD, suggesting the need for careful diagnosis and consideration of BP, even after a diagnosis of ADHD is made (Youngstrom, Arnold, & Frazier, 2010). Likewise, a review of the literature finds evidence for comorbidity rates of CD and BP between 40 – 69% (Kim & Miklowitz, 2002).

Differentiating Unipolar and Bipolar Depression

Overlap in internalizing symptoms, in addition to externalizing symptoms, is a challenge to accurate diagnosis of BP, and differentiating BP from unipolar depression is yet another source of misdiagnosis of BP. Differentiating unipolar depression from BP depression is complex, but research suggests there are meaningful differences in symptom presentation between unipolar and BP depression. BP depression is associated with earlier age of onset and increased severity of symptoms as well as more frequent depressive episodes (Smith et al.,

2011). BP depression is also associated with increased substance comorbidity, family history of manic symptoms, poorer psychosocial functioning and quality of life, as well as conduct problems in childhood and adolescence (Angst et al., 2002; Smith et al., 2011). Research also suggests specific symptoms such as hypersomnia and motor retardation may be more common in BP depression (Bowden, 2005). Improvements in screening methods, such as self-report questionnaires capitalizing on meaningful differences between unipolar and BP depression, can lead to more accurate and earlier diagnosis of BP. Improvements in screening can result in improved prognosis for affected individuals as well as reductions in economic costs and utilization of heath care resources.

Screening Tools for Bipolar Disorders

General Behavior Inventory

Previous research suggests the presence of both manic and depressive dimensions in BP are useful for differentiating BP from other diagnoses (Arrasate et al., 2014). The General Behavior Inventory (GBI) is one measure developed to assess both manic and depressive dimensions in individuals (Depue, Krauss, Spoont, & Arbisi, 1989; Depue et al., 1981). The GBI was developed to identify cases of BP as well as "subsyndromal" cases including BP-NOS/OS-BRD and cyclothymia. Development of the GBI incorporated core behaviors of BP as well as nonbehavioral dimensions (e.g., intensity, duration, frequency, and variability) to create a scale with a low false positive rate for BP. The GBI is a 73-item scale assessing two domains of hypomania/mania and biphasic (i.e., fluctuations) symptoms (28 items), and depression symptoms (46 items) on the basis of intensity, duration, and frequency measured on a 4 point Likert scale (*0* = Never or hardly ever, *3* = Very often almost constantly). Scores on the GBI are broken down into two subscales – Hypomania/Biphasic and Depression Subscales. Internal

consistency of the GBI is excellent ($\alpha = .94$) with strong retest stability (product-moment correlation = .73, 15-week retest interval) (Depue et al., 1981).

The GBI is an effective screening instrument for differentiating between BP or other affective disorders and other psychiatric diagnoses in a variety of populations including both clinical and non-clinical samples. Both Depue et al. (1981) and Mallon, Klein, Bornstein, and Slater (1986) demonstrated 88% accuracy identifying BP and other affective disorders from other diagnoses in adult clinical and non-clinical samples. Depression and Hypomanic/Biphasic Subscales yielded sensitivity and specificity estimates of .78 and .99 (unipolar depression) and .76 and .99 (BP), respectively (Depue et al., 1989). More recent research suggests the GBI is also useful for differentiating BP from ADHD, disruptive behavior disorders, and unipolar depression in adolescents and young adults (Danielson, Youngstrom, Findling, & Calabrese, 2003; Findling et al., 2002; Pendergast et al., 2014).

In addition to self-report, versions of the GBI modified for teacher- (T-GBI) and parent-report (P-GBI) are also available. The P-GBI demonstrates good validity and diagnostic accuracy, however the T-GBI shows more limited clinical utility. The T-GBI was shown to not reliably differentiate BP from non-BP and ADHD cases in youths and to correlate only at a low level on the P-GBI Hypomanic/Biphasic Subscale; correlations with T-GBI, P-GBI, and self-report GBI were close to 0 or negative (Youngstrom, Joseph, & Greene, 2008). However, P-GBI scores led to statistically significant classification of youths across a variety of conditions including comparison of youths with BP from youths with other mood disorders, youths with disruptive behavior disorders, and youths with any mood disorder (Findling et al., 2002; Youngstrom, Findling, Danielson, & Calabrese, 2001; Youngstrom, Genzlinger, Egerton, & Van Meter, 2015). Altogether, research supports the use of the GBI (self-report and P-GBI) as an

effective screening instrument for differentiating affective disorders from other disorders as well as BP from unipolar depression in both child and adult samples.

However, with a total of 73 items, the GBI is a long instrument, and this characteristic may be an impediment to its implementation across various screening environments. Previous research suggests shortened versions of reliable screening instruments may decrease alpha coefficients as a function of less items but with minimal effects on sensitivity and specificity of the original instrument (Shrout & Yager, 1989). Given the high sensitivity and specificity of the GBI with regards to unipolar and BP diagnoses as well as validation across the lifespan and in clinical and non-clinical samples, a shortened version of the GBI warranted investigation. To this end, Youngstrom, Frazier, Demeter, Calabrese, and Findling (2008) developed a ten-item mania rating scale from the P-GBI to screen for BP in children based on the Hypomanic/Biphasic Subscale of the GBI. Efficiency statistics for the shortened ten-item scale (PGBI-10M) were good with an area under the curve (AUC) for differentiating between BP and other diagnoses of .83. The shortened version also discriminated cases of BP significantly better than the original 28-Hypomania/Biphasic Subscale (Youngstrom et al., 2008). However, the PGBI-10M only assesses the hypomanic/biphasic dimension of affective disorders. Successful classification of affective disorders from other disorders and healthy cases as well as differentiation of BP from unipolar depression may be better served with a screening instrument that assesses depression and hypomanic/biphasic dimensions.

7 Up 7 Down Inventory

The 7 Up 7 Down Inventory (7U7D) is a 14-item scale carved from the GBI that captures both hypomanic/biphasic as well as depression dimensions (Youngstrom, Murray, Johnson, & Findling, 2013). Items from the GBI were chosen based on exploratory factor analyses to identify top-ranked items from the Depression and Hypomania/Biphasic Subscales utilizing data

from adult and youth samples as well as clinical and non-clinical samples. The resulting carved scale of 7 items from the Hypomania/Biphasic Subscale and 7 items from the Depression Subscale were chosen based on item ranking from exploratory factor analysis and adding the necessary number of items to meet internal reliability criterion of >.7. The resulting Hypomanic/Biphasic Subscale (7U) and Depression Subscale (7D) comprising the 7U7D have internal reliabilities of .81 - .93 in youth samples and correlations of .85 - .92 with original full-length GBI subscales. Additionally, 7U7D AUCs for the 7U and 7D subscales range from .59 - .67, respectively, and are statistically similar to AUC estimates for full-length GBI subscales (Youngstrom et al., 2013).

The 7U7D is a promising self-report measure for identifying BP and discriminating between BP and unipolar depression in screening environments due to its brevity, inclusion of both depressive and hypomanic/biphasic dimensions, and psychometric properties similar to the original GBI. However, to date, there are no published recommended cut scores or diagnostic likelihood ratios (DiLRs) for the 7U7D needed for clinical decision-making to improve diagnostic accuracy in differentiating between BP and other diagnoses.

Study Aims

The first aim of the current study is to re-evaluate sensitivity and specificity through Receiver Operating Characteristic (ROC) analyses in order to establish diagnostic efficiency for 7U7D subscales in differentiating BP from other diagnoses as well as BP from unipolar depression in an outpatient youth sample utilizing self- and caregiver-report. A second aim of the current study is to calculate optimal cut scores for each subscale of the 7U7D and present DiLRs to be used in a two-step fashion to first aid clinicians in identifying the presence of mood disorders from other disorders and second to separate BP from unipolar depression. A third aim of the current study is to explore the clinical utility of using 7U7D recommended cut scores and

DiLRs with a nomogram method to determine the posterior probability of accurately diagnosing BP in a youth outpatient example (Jenkins, Youngstrom, Washburn, & Youngstrom, 2011). Exploratory analyses will utilize multinomial logistic regressions to predict individual membership in one of three diagnostic groups (i.e., unipolar, bipolar, or no mood disorder) based on 7U7D scores in a one-step fashion.

METHODS

Participants

Youths 5 to 18 years of age and youth caregivers were recruited from outpatient mental health centers. Eligibility requirements for both studies required youths and caregivers to be fluent in English. Youths with a diagnosis of pervasive developmental disorder or IQ < 70 were excluded from both studies.

The first sample ($N = 909$) was recruited from an academic mental health clinic located within an urban university psychiatry department in Cleveland, Ohio. Families were referred to the clinic from within the psychiatry department (clinical research center or pediatric psychopharmacology clinical trials) or from outside referrals as well as from advertisements, and referrals from within the community (Findling et al., 2002; Findling et al., 2001). The clinical research center recruited youths seen within the psychiatry department with a BP diagnosis, at high risk for BP (caregiver seen at adult mood disorder clinic), or youths without a psychiatric diagnosis. Psychopharmacology clinical trials within the psychiatry department recruited youths with a range of psychiatric diagnoses. Caregivers completed a telephone screen before meeting with research assistants in the clinic to complete a semi-structured diagnostic interview.

The second sample ($N = 828$) was recruited from several urban community mental health centers in Cleveland, Ohio (Youngstrom et al., 2005). A random subset of families seeking outpatient care for their youths for a variety of psychiatric concerns were asked to take part in the study.

Institutional Review Boards at respective treatment sites approved all study procedures. Assent and consent were obtained from caregivers and youths prior to participation. Diagnostic interviews were completed sequentially but separately with youths and caregivers. Caregivers completed the P-GBI regarding youths and self-report scores on the GBI were obtained from youths 11 years or older. Families were compensated for participation.

Measures

Schedule for Affective Disorders and Schizophrenia for School-Age Children Present and Lifetime Version (K-SADS-PL; Kaufman et al., 1997) The K-SADS is a well-validated, semi-structured diagnostic interview for assessing the presence of current and lifetime psychiatric diagnoses in youths. K-SADS interviews were completed with youths and caregivers separately by highly trained raters. Raters were blind to GBI scores. The Longitudinal Evaluation of All Available Data (LEAD) standard for diagnosing psychiatric disorders was used to take into account information from all available sources (Spitzer, 1983). For this study, information from the caregiver K-SADS interview, youth K-SADS interview, family history, and clinical judgment were all considered when making diagnostic decisions. Additionally, all diagnoses underwent a consensus review process whereby a licensed psychologist or psychiatrist reviewed and confirmed the presence of all diagnoses according to Diagnostics and Statistics Manual – 4th Edition (DSM-IV) criteria (American Psychiatric Association, 2000). Three different versions of the K-SADS were used across the two samples of families. The K-SADS-Epidemiologic version (K-SADS-E; Orvaschel, 1994), which contains specific questions about suicidal behaviors, was conducted with the first 200 families recruited from academic mental health clinics. A second version of the K-SADS, the K-SADS- Present and Lifetime Version (K-SADS-PL; Kaufman et al., 1997) was conducted with the majority sample of families recruited from academic mental health clinics. A third version of the K-SADS, the K-SADS-PL with additional mood modules

from Washington University (WASH-U-KSADS; Geller et al., 2001) was conducted with all families recruited from community mental health centers as well as around 175 cases from the academic mental health clinic.

General Behavior Inventory (GBI; Depue et al., 1981). The GBI was administered as a self-report measure to youths as well as caregiver informants. The GBI was scored according to the Likert type scoring method described by Depue et al. (1981) with item scores ranging from 0 to 3.

7 Up 7 Down Inventory (7U7D; Youngstrom et al., 2013). 7U7D scores were calculated for youth self-report and caregiver-report from full-length GBIs. Items 22, 30, 31, 38, 43, 46, and 64 from the original GBI were summed to calculate the 7U Subscale. Items 23, 34, 47, 56, 62, 63, and 73 were summed to calculate the 7D Subscale. Scores were prorated if respondents skipped 1-2 items on the subscales but answered at least 5 of the items. Therefore, non-integer scores are observed in the analyses and recommended cut points.

Analysis Plan

All analyses were done using *R* version 3.3.2 and the pROC package (Robin et al., 2011). Efficiency of 7U and 7D subscales to identify mood disorders broadly and BP disorders utilizing self-report GBI and P-GBI will be assessed separately utilizing ROC analyses. Descriptive analyses will determine any meaningful group differences between community clinic outpatient and academic clinic outpatient samples that may impact efficiency estimates (e.g., gender and age) prior to analyses.

ROC analyses are useful for clinical decision-making because they produce AUCs that provide accuracy of a given measure for identifying a dichotomous outcome (e.g., mood disorder diagnosis vs. no mood disorder diagnosis) taking into consideration the relative frequencies of true positive, false positive, true negative, and false negative cases (Metz, 1978). ROC curves

allow clinicians to make decisions about cut points based on the risks and benefits of a test being sensitive (i.e., proportion of true positive cases) versus specific (i.e., proportion of true negative cases). AUCs around .50 are considered small (around chance), .60 considered medium, .70 large, and .80 or above exceptional for diagnostic accuracy (Swets, Dawes, & Monahan, 2000). Plots of ROC curves are also useful graphical tools for making decisions regarding optimal cut points and comparing the relative performance of two measures; comparisons of two ROC curves and decision for optimal cut points are accomplished through quantitative comparisons (Zweig & Campbell, 1993). Optimal cut scores for the present analyses were identified using Youden's *J* statistic which yields a cut-off point that maximizes the sensitivity and specificity of the measure, or the threshold point that maximizes the distance from the diagonal reference line (Robin et al., 2011). Youngstrom (2014) provides a primer on ROC analyses in the context of clinical decision-making.

ROC analyses will be conducted in a two-step fashion (Aims 1 and 2). Initial ROC analyses will be conducted to estimate efficiency for identifying individuals with any mood disorder from other diagnoses utilizing the 7D. A second ROC analysis will be performed on the subset of youths identified as "high risk" for any mood disorder (i.e., 7D scores above the identified optimal cut point for presence of a mood disorder) utilizing the 7U to estimate efficiency of differentiating BP from other mood disorders. A parallel second ROC analysis will also be performed on youths identified as "low risk" for a mood disorder (i.e., 7D scores below the identified optimal cut point for presence of a mood disorder) utilizing the 7U to distinguish cases of BP without a history of depression (see Figure 1) from other diagnoses. The two-step ROC analysis will account for heterogeneous current and past symptom presentations that can lead to a diagnosis of BP (see Figure 2). The two-step ROC analysis allows for use of both the

7U to capture present and/or past hypomania/mania symptoms as well as the 7D to capture present and/or past depression symptoms.

Diagnostic accuracy of each subscale will also be compared between caregiver- and self-report. Comparison of AUCs will be done using a nonparametric approach devised by Venkatraman (2000), which accounts for implicit correlations between AUCs derived from the same sample. Venkatraman's test compares ROC curves at all points rather than just the overall numerical AUC estimate, which allows for detection of significant differences in efficiency despite similar overall AUC estimates. DiLRs for optimal 7U7D subscale cut points will be calculated to determine increases in odds associated with different cut points for the presence of any mood disorder (step 1) and BP (step 2). Utility of two-step optimal cut scores from the 7U7D will be presented in a clinical example using the nomogram method. The nomogram method combines the pretest probability of a diagnosis (i.e., base rates for diagnoses) and the likelihood ratio corresponding to an individual's score on a given measure to produce a posttest probability for a target diagnosis.

Exploratory analyses utilizing a multinomial logistic regression will explore the probability of being diagnosed with BP, unipolar depression, or other disorder (including no disorder) based on subscale scores of the 7U7D. The purpose of the multinomial logistic regression is to explore the diagnostic accuracy and clinical utility of using the 7U7D as a one-step, rather than two-step, approach in differentiating youth BP from other diagnoses.

RESULTS

Demographics and Preliminary Analyses

Prior to combining the academic mental health outpatient sample and the community mental health outpatient sample, demographic characteristics were compared between the two groups (see Table 1). The academic mental health outpatient sample had a significantly higher proportion ($X^2(1) = 549.55$, $p < .01$) of Caucasian youths (79%) compared with the community sample (22%). Youths in the academic sample were also significantly older ($t(1669) = 2.47$, $p = .01$; $M = 11.31$, $SD = 3.39$) than the community sample ($M = 10.90$, $SD = 3.42$) but had less comorbid Axis I diagnoses ($t(1662) = 9.36$, $p = .01$; $M = 2.05$, $SD = 1.24$) compared with the community sample ($M = 2.65$, $SD = 1.37$). As expected based on referral patterns, the academic outpatient sample had a significantly higher proportion ($X^2(1) = 196.11$, $p < .01$) of BP youths (51%) compared with the community sample (18%). All significant differences between the samples were small in effect size with all *d*s less than or close to 0.30, with the exception of effect size for race ($d = .57$). The large effect size for race reflects notable racial and ethnic differences between the samples. The academic outpatient sample had a significantly higher proportion of Caucasian youths compared with the community outpatient sample. There were no significant sex differences. Given significant sample differences in demographics and clinical characteristics between the academic and community samples, demographic and clinical characteristics were included in models examining performance of the 7U7D.

Table 2 compares demographic and clinical characteristics of the pooled sample between youths with a BP diagnosis and youths without BP. BP youths had significantly more comorbid

Axis I diagnoses, higher rates of comorbid ADHD and ODD, were less likely to be Caucasian, and had significantly higher scores on the 7U7D subscales according to both caregiver- and self-report (all *p*s <.05). There were no significant group differences on sex, age, or rate of comorbid anxiety.

Internal consistencies of 7U7D subscales were calculated. Alpha coefficients from the caregiver-report for the 7U and 7D were .82 and .90, respectively, and .79 and .90, from self-report. Internal consistencies in the present sample are similar to other samples (Mesman, Youngstrom, Juliana, Nolen, & Hillegers, 2017; Youngstrom et al., 2013).

Diagnostic Efficiency

Table 3 presents AUC values for 7U7D subscales split by caregiver and self-report. AUCs for 7D and 7U caregiver-report subscales predicting any mood disorder diagnoses were .81, p <.01, 95% CI [.79 to .83] and .65, p <.01, 95% CI [.62 to .68], respectively. Caregiver-report AUCs for 7D and 7U subscales predicting bipolar disorders were .70, p <.01, 95% CI [.67 to .73] and .76, p <.01, 95% CI [.73 to .78], respectively. Diagnostic efficiency of the caregiver-report 7U subscale predicting bipolar disorders applied to samples identified as either "low risk " (i.e., scores less than optimal cut score of 3.25 on the 7D) or "high risk" (i.e., scores equal to or greater than the optimal cut score of 3.25 on the 7D) yielded AUC estimates of .72, p <.01, 95% CI [.67 to .77] in the low risk sample and .72 p <.01, 95% CI [.68 to .76] in the high risk sample.

Overall, AUCs for self-reported 7U7D subscales were significantly lower compared with caregiver-report (all Venkatraman's test *p*s <.01). AUCs for 7D and 7U self-report subscale predicting any mood disorder diagnoses were .67, p <.01, 95% CI [.63 to .71] and .58, p <.01, 95% CI [.54, .62], respectively. Self-report AUCs for 7D and 7U predicting bipolar disorders were .56, p < .05, 95% CI [.52 to .61] and .60, p <.01, 95% CI [.56 to .64], respectively. Diagnostic efficiency of the self-report 7U subscale predicting BP applied to samples identified

as either "low risk " (i.e., scores less than 2.18 on the 7D) or "high risk" (i.e., scores equal to or greater than 2.18 on the 7D) yielded AUC estimates of .55, $p > .05$, 95% CI [.45 to .64] in the low risk sample and .60, $p < .01$, 95% CI [.54 to .65] in the high risk sample. Figure 3 compares AUC curves for 7U7D subscales predicting target diagnoses by informant report.

Incremental Validity. Logistic regressions predicting diagnoses of any mood disorders broadly or BP tested the incremental validity of 7U7D subscales above and beyond demographic and clinical variables. Model 1 controlled for demographic and clinical variables including child sex, age of child, number of child Axis I comorbidities, and race. Model 2 added both caregiver- and self-reported 7U7D subscales. Model 3 added interaction terms between demographic variables and 7U7D subscales, testing whether demographic variables affected scale accuracy. Comparison of model fit determined if subsequent models were an improvement over previous models.

Any Mood Disorders. Model 1 explained 30% (Nagelkerke's $R^2 = .299$) of variance in broad mood disorder status. Mood disorder status was significantly associated with being female ($B = 0.58$, $SE = 0.12$, $z = 4.81$, $p < .001$), older age ($B = 0.14$, $SE = 0.02$, $z = 8.11$, $p < .001$), higher number of comorbidities ($B = 0.82$, $SE = 0.06$, $z = 14.69$, $p < .001$), and being white ($B = 1.21$, $SE = 0.12$, z = 10.14, $p < .001$). Comparison of model fit revealed Model 2 was a significant improvement over Model 1 (ΔX^2 (4) = 111.64, $p < .001$) and explained an additional 18% of variance in broad mood disorder status (ΔNagelkerke's $R^2 = .18$). Overall, Model 2 explained 48% (Nagelkerke's $R^2 = .482$) of variance in mood disorder status. All predictors from Model 1 remained significant with the exception of child race. In addition to demographic predictors, higher scores on the caregiver-reported 7D ($B = 0.32$, $SE = 0.04$, z = 8.83, $p < .01$), but not caregiver-reported 7U or self-reported subscales, were significantly associated with mood disorder status. Comparison of model fit confirmed Model 3 was not a significant improvement

over Model 2 (ΔX^2 (12) = 20.79, p = .05). Overall, Model 3 explained, 51% (Nagelkerke's R^2 = .508) of the variance in broad mood disorder status, which only reflects an additional 3% (ΔNagelkerke's R^2 = .03) of explained variance in broad mood disorder status. The only significant interaction term related to broad mood disorder status was between caregiver-reported 7U and child race (B = 0.16, SE = 0.07, z = 2.35, p = .02).

Bipolar Disorders. Model 1 explained 22% (Nagelkerke's R^2 = .217) of variance in BP status. BP status was significantly associated with number of Axis I comorbidities (B = 0.55, SE = 0.05, z = 11.98, p <.001) and being white (B = 1.48, SE = 0.12, z = 12.15, p <.001). Comparison of model fit revealed Model 2 was a significant improvement over (Model 1, ΔX^2 (4) = 118.06, p <.001) and explained an additional 9% of variance in broad mood disorder status (ΔNagelkerke's R^2 = .09). Overall, Model 2 explained 31% (Nagelkerke's R^2 = .308) of variance in BP status. In addition to significant predictors from Model 1, higher scores on the caregiver-reported 7D (B = 0.06, SE = 0.02, z = 2.55, p =.01) and 7U *(*B = 0.19, SE = 0.03, z = 7.08, p <.001)*,* but not self-reported subscales, were significantly associated with BP status. Comparison of model fit confirmed Model 3 was not a significant improvement over Model 2 (ΔX^2 (12) = 15.52, p = .21). Overall, Model 3 explained, 33% (Nagelkerke's R^2 = .33) of the variance in BP status, which only reflects an additional 2% (ΔNagelkerke's R^2 = .02) of explained variance in BP status. The only significant interaction term related to BP status was between caregiver-reported 7U and child race (B = 0.15, SE = 0.07, z = 2.63, p < .01).

Diagnostic Likelihood Ratios

Diagnostic likelihood ratios were calculated separately for caregiver- and self-report 7U7D subscales and are presented in Table 4. Given that models including interaction terms with 7U7D subscales were not a significant improvement, DiLRs were not calculated separately for

child sex, age, or race, but rather based on informant report source. DiLRs were calculated using a two-step process.

Using optimal cut scores for the first step, a caregiver 7D score of 3.26 or higher (Sensitivity = 71%; Specificity = 78%) resulted in 3.26 increase in odds of a broad mood disorder diagnosis, whereas a caregiver 7D score below 3.26 resulted in a reduced likelihood of a broad mood disorder diagnosis (DiLR = 0.38). Optimal cut scores were slightly lower for the self-report 7D and resulted in lower DiLRs. A self-report 7D of 2.18 or higher (Sensitivity = 74%; Specificity = 52%) resulted in a 1.55 increase in odds of a broad mood disorder diagnosis whereas a self-report 7D scores below 2.18 resulted in reduced likelihood of a broad mood disorder diagnosis (DiLR = 0.49).

Optimal cut scores were next calculated when applying the 7U to predict BP diagnosis in the *low risk* sample. A caregiver 7U score of 3.26 or higher (Sensitivity = 66%; Specificity = 68%) resulted in a 2.11 increase in odds of a BP diagnosis whereas a caregiver 7U score below 3.26 resulted in a reduced likelihood in odds of a BP diagnosis (DiLR = 0.49). Optimal cut scores were again slightly lower in the self-report 7U and resulted in lower DiLRs. A self-report 7U of 2.68 or higher (Sensitivity = 64%; Specificity = 50%) resulted in a 1.27 increase in odds of a BP whereas a self-report 7U below 2.68 resulted in a decrease in odds of a bipolar disorder diagnosis (DiLR = 0.73).

Optimal cut scores were also calculated when applying the 7U to predict bipolar disorder diagnoses in the *high risk* sample. A caregiver 7U score of 5.43 or higher (Sensitivity = 70%; Specificity 64%) resulted in a 1.94 increase in odds of a BP diagnosis whereas a caregiver 7U score below 5.43 resulted in a reduced likelihood in odds of a BP diagnosis (DiLR = 0.47). A self-report 7U score of 5.31 or higher (Sensitivity = 71%; Specificity = 44%) was not associated

with a change in odds of a bipolar disorder (DiLR = 1.00) whereas a self-report 7U below 5.31 was associated with a decrease in odds of a BP diagnosis (DiLR = 0.51).

Exploratory Analysis

Exploratory multinomial logistic regressions were conducted to determine performance of the 7U7D using a one-step rather than two-step process for predicting a diagnosis of BP or unipolar mood disorder compared with other diagnoses. The dependent diagnostic variable was split into three groups based on primary DSM-IV diagnosis. Youths with any BP diagnosis were coded as one group, any unipolar mood diagnosis was coded as another group, and youths without a BP or unipolar mood diagnosis made up the third reference group (i.e., "clinical sample"). Two models were run for each informant group. In the first model, youth demographics and clinical characteristics were used as predictors (i.e., sex, race, age, number of Axis 1 diagnoses). The second model included 7U7D subscales as predictors in addition to demographic and clinical characteristics. Model chi-square statistics and changes in pseudo R^2s compared model fits. Model 2 fit significantly better than model 1 in the caregiver-report sample (ΔX^2 (4) = 420, p <.01) with a McFadden R^2 = .29 (Δ McFadden R^2 from Model 1 = .135) suggesting excellent model fit (McFadden, 1977). Model 2 also fit significantly better than model 1 in the caregiver-report sample (ΔX^2 (4) = 42, p <.01) with a McFadden R^2 = .13 (Δ McFadden R^2 from Model 1 = .03). McFadden (1977) cautions that the McFadden pseudo R^2 is different than the traditional R^2 for maximum likelihood estimation and yields considerably lower values. Results of the caregiver-report and self-report multinomial logistic regressions with all predictors (i.e., model 2) are presented in Tables 5 and 6, respectively.

Unipolar diagnosis compared to the clinical sample. Results from the caregiver-report suggest that sex (*OR* = 1.74 [95% CI 1.26 to 2.40]), age (*OR* = 1.16 [95% CI 1.11 to 1.22]), and number of Axis I diagnoses (*OR* = 1.94 [95% CI 1.69 to 2.24]) were all significant predictors (*p*s

<.01). Race was not a significant predictor. Being female, older, and having more Axis I diagnoses increase the probability of having a unipolar mood diagnosis compared with another primary Axis I diagnosis or no diagnosis. When holding demographic and clinical variables constant, scores on the 7D predicted unipolar mood diagnosis (*OR* = 1.34 [95% CI 1.28 to 1.40], *p* <.01). Each point increase on the 7D caregiver report is associated with a 1.34 change in odds for unipolar mood disorder diagnosis. The 7U also significantly predicted unipolar mood diagnosis (*OR* = 0.90 [95% CI 0.86 to 0.94], *p* <.01). Each point increase on the 7U caregiver report is associated with a 0.90 change in odds for unipolar mood disorder diagnosis; higher scores on the 7U decrease the probability of a unipolar mood diagnosis.

Results from self-report 7U7D suggest that race (*OR* = 1.49 [95% CI 1.01 to 2.19], *p* <.05), age (*OR* = 1.28 [95% CI 1.15 to 1.42], *p* <.01), and number of Axis I diagnoses (*OR* = 1.81 [95% CI 1.53 to 2.14], *p* <.01) were all significant predictors. Unlike caregiver-report, sex was not a significant predictor. Being Caucasian, older, and having more Axis I diagnoses increase the probability of having a unipolar mood diagnosis compared with another primary Axis I diagnosis or no diagnosis. When holding demographic and clinical variables constant, self-report scores on the 7D predicted unipolar mood diagnosis (*OR* = 1.12 [95% CI 1.07 to 1.17], *p* <.01). Each point increase on the 7D caregiver report is associated with a 1.12 change in odds for unipolar mood disorder diagnosis. Unlike caregiver-report, scores on the 7U self-report were not a significant predictor of unipolar mood diagnosis.

Bipolar diagnosis compared to the clinical sample. Results from caregiver-report 7U7D suggest that sex (*OR* = 1.58 [95% CI 1.13 to 2.21]), race (*OR* = 3.98 [95% CI 2.85 to 5.56]), and number of Axis I diagnoses (*OR* = 2.18 [95% CI 1.89 to 2.52]) were all significant predictors (*p*s <.01). Age was not a significant predictor. Being female, Caucasian, and having more Axis I diagnoses increase the probability of having a BP diagnosis compared with another primary Axis

I diagnoses or no diagnosis. When holding demographic and clinical variables constant, scores on the 7D predicted BP diagnosis (*OR* = 1.27 [95% CI 1.21 to 1.32], *p* <.01). Each point increase on the 7D caregiver report is associated with a 1.27 change in odds for BP diagnosis. The 7U also significantly predicted BP diagnosis (*OR* = 1.14 [95% CI 1.09 to 1.19], *p* <.01). Each point increase on the 7U caregiver report is associated with a 1.14 change in odds for BP diagnosis.

Results from 7U7D self-report suggest that sex (*OR* = 1.95 [95% CI 1.26 to 3.01) race (*OR* = 3.73 [95% CI 2.43 to 5.71]), age (*OR* = 1.30 [95% CI 1.16 to 1.48]), and number of Axis I diagnoses (*OR* = 2.09 [95% CI 1.75 to 2.50]) were all significant predictors (*p*s <.01). Unlike caregiver-report, age was not a significant predictor in the self-report models. Being female, Caucasian, older, and having more Axis I diagnoses increase the probability of having a BP diagnosis compared with another primary Axis I diagnosis or no diagnosis. When holding demographic and clinical variables constant, scores on the 7D predicted BP diagnosis (*OR* = 1.06 [95% CI 1.01 to 1.11], *p* <.05). Each point increase on the 7D caregiver report is associated with a 1.06 change in odds for BP diagnosis. Unlike caregiver-report, scores on the 7U self-report were not a significant predictor of BP diagnosis.

DISCUSSION

The overall aim of the present study was to examine the diagnostic efficiency of caregiver- and self-report versions of the 7U7D in discriminating youth BP from unipolar depression and other diagnoses in an outpatient clinical setting. Overall, both caregiver-and self-reported 7U7D subscales significantly differentiated broad mood disorder as well as BP diagnoses compared with other diagnoses. AUCs for both 7U and 7D subscales differentiating broad mood disorder and BP diagnoses compared with other DSM-IV diagnoses produced medium to large effects and are similar to previously published 7U7D AUC estimates (Swets, Dawes, & Monahan, 2000; Youngstrom et al., 2013). These medium to large effects suggest the 7U7D is a clinically useful tool for differentiating youth mood disorders from other diagnoses. This finding is especially pertinent given the heterogeneous nature of BP and symptom overlap between BP and other diagnoses (e.g., unipolar depression), which contributes to high rates of misdiagnosis and appropriate treatment delay in youth BP (Biederman, 1998; Kim & Miklowitz, 2002; Marchand, 2006; Youngstrom et al., 2010).

Caregiver-reported 7U7D subscales performed significantly better than youth self-report in differentiating any mood and BP diagnoses from other primary Axis I diagnoses. This finding is consistent with the results of previous studies showing caregiver-report performs consistently better than youth self-report differentiating mood disorders and BP from other diagnoses (Youngstrom et al., 2001; Youngstrom et al., 2015). These findings suggest that, although youth self-report can generate clinically useful information, clinicians should try to collect caregiver-

report when possible to improve accuracy in making diagnostic decisions and planning effective treatments.

Examination of the incremental validity of 7U7D subscales predicting mood disorders broadly strengthens support for gathering caregiver-report. Only the caregiver-reported 7D significantly predicted any mood disorder status after controlling for demographics and number of comorbidities while both caregiver-reported 7D *and* 7U subscales predicted BP status after controlling for demographic and clinical variables. Self-reported 7U7D did not significantly predict either diagnostic category after controlling for demographic and clinical variables. Age, race, and sex by caregiver- and self-report 7U7D interactions did not significantly improve variance explained in broad mood disorder or BP diagnosis status. Since interaction terms did not significantly improve variance in diagnostic classification, DiLRs were calculated separately for informant source only and not based on demographic characteristics.

The second aim of the present study was to calculate optimal cut scores for 7U7D and to present DiLRs to be used in a two-step fashion to first aid clinicians in identifying the presence of mood disorders from other disorders and secondly to separate BP from unipolar depression. To date, no studies have published DiLRs for the 7U7D. A two-step DiLR process uses both the 7U to capture present and/or past hypomania/mania symptoms and the 7D to capture present and or/past depression symptoms. Using both subscales addresses the heterogeneous nature of BP and previous research suggesting both manic and depressive dimensions are useful for identifying BP (Arrasate et al., 2014). The two-step DiLR process is also in line with two-step DiLRs published for the GBI, from which the 7U7D is carved (Pendergast et al., 2014). Further support for the two-step DiLR process comes from exploratory analyses examining 7U7D in a one-step fashion which showed subscale scores differentially influenced odds ratios depending on diagnostic group; higher scores on the 7D increase odds for both unipolar depression and BP

diagnosis while higher 7U scores decrease the odds of a unipolar depression and increase the odds of a BP diagnosis.

DiLR values presented in Table 4 show the 7U7D contributes clinically meaningful information in differentiating broad mood disorders from other clinical diagnoses as well as BP from unipolar mood disorders. DiLRs presented show that categorizing 7D scores into low and high thresholds change the odds of any mood disorder diagnosis from one and half to over threefold depending on informant report. Additionally, 7U high threshold scores change the odds of differentiating BP diagnosis from other diagnoses to as much as double depending on previous 7D scores and informant report.

DiLRs make it easier for clinicians to interpret scores on report measures within the context of other clinically relevant information (e.g., base rates of diagnoses in clinics, family history) to produce better estimates of risk utilizing the nomogram method commonly used in evidence-based medicine (Jenkins, Youngstrom, Washburn, & Youngstrom, 2011; Youngstrom, 2014). A third aim of the current study was to explore the clinical utility of using 7U7D DiLRs with a nomogram method. The following clinical vignette illustrates the application of incorporating base rates and DiLRs from the 7U7D to guide clinical decision-making.

Clinical Vignette

A 14-year old boy, Alex, is referred to your community outpatient clinic by his mother who is concerned about his poor school attendance, social withdrawal, sleep problems, and frequent irritability. Alex's mother reports she first noticed these behaviors when Alex was around 12 years old but that these behaviors have been increasing in severity and frequency within the past 6 months. You have trouble-engaging Alex during this initial appointment but are able to complete a brief background interview with Alex's mother and she also fills out the 7U7D. In speaking with Alex's mother, you learn Alex's father was diagnosed with a mood

disorder, although Alex's mother is unsure of the specific diagnosis but your case conceptualization now includes mood disorder as a likely diagnosis. You are unsure of the specific base rate of mood disorders in your clinic so you use the national base rate for youth mood disorders of 14.3% as your starting point and pretest probability (Merikangas, He, Burstein, Swanson, & Avenevoli, 2010). You see that the results from Alex's mother's 7U7D shows a score of 8 on the 7D which corresponds with a DiLR of 3.26 and places Alex in the "high risk" category according to the 7D. You draw a line from Alex's pretest probability through the likelihood ratio of 3.26 (Figure 3) and see that Alex's post-test probability for any mood disorder is now just over 30%. To incorporate the 7U score, move the post-test probability from the first nomogram to the pre-test probability in the second nomogram (now the starting probability) and draw a line from this probability through the likelihood ratio of 1.94 which corresponds to Alex's mother's rating of a 6 on the 7U. The addition of the 7U score now raises the posterior probability to around 50% for BP (see Figure 3) which places Alex in the "Yellow Zone" between test and treat thresholds (Youngstrom, 2014). Without further assessment, you are able to incorporate both depression *and* hypomanic/manic dimensions of Alex's current symptoms using the nomogram method, which guides you towards an evidence-based decision that a low risk treatment, like psychotherapy, targeting BP is the best course of action.

Limitations

The present study has several limitations, which should be taken into account when interpreting results. In an effort to increase sample size and robustness of AUC and DiLR estimates, samples reflecting two different outpatient settings (i.e., academic mental health center and several community mental health centers), referral patterns, and different versions of the K-SADS diagnostic interview were combined to create the present study sample. Despite similar training practices and overlapping study staff across sites, differences may exist with regards to

diagnostic ratings. Demographic and clinical differences, albeit mostly small in effect size, did exist between study sites and may influence the findings.

Secondly, self-report 7U7D was only available for youths 11 years of age or older. Given the relatively high reading level and length (i.e., 73-items) of the GBI from which the 7U7D items were carved, self-report may not be appropriate for youths younger than 11 years of age. Comparisons in the current study between youth and caregiver-report are limited in age range. Future studies should investigate the clinical utility and feasibility of administering the much shorter self-report 7U7D in a younger population.

Thirdly, although models including interactions between child and caregiver characteristics with the 7U7D were not statistically significant overall, there were significant individual interaction terms between caregiver-reported 7U and child race predicting broad mood disorder and BP status. These results suggest cultural factors (e.g., race and ethnicity) may influence performance of the 7U7D. Future studies should investigate performance of the 7U7D with a focus on the impact of specific cultural factors on diagnostic accuracy of the 7U7D.

Finally, diagnoses in the present study are based on DSM-IV diagnostic criteria. Although changes to mood disorder criteria were relatively minor in the DSM-5, future research should confirm whether or not changes in diagnostic criteria affect diagnostic accuracy of the 7U7D.

Conclusions

While both caregiver and self-reported 7U7D subscales are useful for differentiating the presence of broad mood disorders as well as BP from other diagnoses, caregiver-report demonstrated significantly better efficiency differentiating diagnoses. The 7U7D is a promising tool for differentiating broad mood disorders and BP in a clinical setting given its brevity, medium to large effects differentiating diagnoses, and its assessment of both hypomanic/manic

and depression dimensions. Presentation of DiLRs for both caregiver and self-report increases the clinical utility of the 7U7D by providing clinicians with accessible evidence-based estimates to guide decisions pertaining to follow-up assessments and course of treatment.

Table 1
Demographics and Clinical Characteristics by Clinic Sample

Variable	*Academic Clinic* (*N* = 907)	*Community Clinic* (*N* = 828)	*Test Statistic*	*p*	*Effect Size*[b]
Male, % (*n*)	63% (565)	60% (496)	X^2 (1) = 1.97	.16	0.04
Age, M (SD)	11.31 (3.39)	10.90 (3.42)	*t*(1669) = 2.47	.01	0.06
Caucasian, % (*n*)	79% (702)	22% (185)	X^2 (1) = 549.55	<.01	0.57
Number Axis I Diagnoses	2.05 (1.24)	2.65 (1.37)	*t*(1662) = 9.36[a]	.01	0.22
Any BP Diagnosis, % (*n*)	51% (451)	18% (153)	X^2 (1) = 196.11	<.01	0.34
Any Unipolar Mood Diagnosis, % (*n)*	19% (172)	28% (230)	X^2 (1) = 17.87	<.01	0.10
Any Anxiety Diagnosis, % (*n*)	10% (93)	26% (213)	X^2 (1) = 70.52	<.01	0.20
Any ADHD Diagnosis, % (*n*)	58% (515)	63% (520)	X^2 (1) = 4.81	.03	0.05
Any ODD Diagnosis, % (*n*)	29% (255)	39% (321)	X^2 (1) = 20.75	<.01	0.11
7d – Caregiver Report	6.24 (5.62)	3.91 (4.29)	*t*(1141) = 9.00[a]	<.01	0.23
7u – Caregiver Report	5.30 (4.74)	4.33 (3.95)	*t*(1217) = 4.17[a]	<.01	0.11
7d – Self Report	6.54 (6.09)	5.74 (5.22)	*t*(552) = 1.86[a]	.06	0.07
7u – Self Report	5.60 (4.25)	6.01 (4.32)	*t*(761) = 1.29	.20	0.05

Note: Where data points were missing, effect sizes were calculated out of total number of available cases.
[a]Equal variances not assumed, Levene's test $p<.05$. [b]Effect sizes are Cohen's *d* (means) or phi (proportions).

Table 2
Demographics and Clinical Characteristics by Diagnosis (Whole Sample)

Variable	*Any BP* (*n* = 604)	*No BP* (*n* = 1112)	*Test statistic*	*p*	*Effect Size*[b]
Male, % (*n*)	62% (373)	61% (682)	X^2 (1) = 0.03	.86	0.00
Age, m (SD)	10.98 (3.50)	11.17 (3.36)	*t*(1660) = 1.05	.29	0.03
Caucasian, % (*n*)	42% (422)	70% (459)	X^2 (1) = 125.86	<.01	0.27
Number Axis I Diagnoses	2.83 (1.30)	2.07 (1.29)	*t*(1710) = 11.6	<.01	0.27
Any Unipolar Mood Diagnosis, % (*n)*	0% (0)	36% (400)	X^2 (1) = 284.22	<.01	0.41
Any Anxiety Diagnosis, % (*n*)	17% (105)	18% (200)	X^2 (1) = 0.12	.73	0.01
Any ADHD Diagnosis, % (*n*)	67% (407)	57% (628)	X^2 (1) = 18.56	<.01	0.10
Any ODD Diagnosis, % (*n*)	38% (229)	31% (347)	X^2 (1) = 7.54	<.01	0.07
7D – Caregiver Report	7.23 (5.36)	3.98 (4.59)	*t*(729) = 11.00[a]	<.01	0.28
7U – Caregiver Report	7.51 (4.66)	3.55 (3.58)	*t*(674) = 15.90[a]	<.01	0.39
7D – Self Report	7.01 (5.95)	5.68 (5.41)	*t*(353) = 2.84[a]	<.01	0.10
7U – Self Report	6.98 (4.58)	5.42 (4.10)	*t*(760) = 4.57	<.01	0.16

Note: Where data points were missing, effect sizes were calculated out of total number of available cases.
[a]Equal variances not assumed, Levene's test $p<.05$. [b]Effect sizes are Cohen's *d* (means) or phi (proportions).

Table 3
AUC estimates by informant report and target diagnosis.

	AUC [95% CI]			
Subscale	*Any Mood vs. All*	*Bipolar vs. All*	*Bipolar vs. All (Low Score 7D)*	*Bipolar vs. All (High Score 7D)*
7D – Caregiver Report[a]	.81** [.79, .83]	.70** [.67, .73]	-	-
7U – Caregiver Report[a]	.65** [.62, .68]	.76** [.73, .78]	.72** [.67, .77]	.72** [.68, .76]
7D – Self Report	.67** [.63, .71]	.56* [.52, .61]	-	-
7U – Self Report	.58** [.54, .62]	.60** [.56, .64]	.55 [.45, .64]	.60** [.54, .65]

[a]Area Under the Curve (AUC) values significantly better in caregiver-report sample compared with self-report sample, Venkatraman's test all *ps* <.05.
*p<.05, **p<.01

Table 4
DiLRs based on two-step process by informant report.

	Low Optimal Threshold		High Optimal Threshold		
	Range	DiLR	Range	DiLR	Sensitivity, Specificity
Step one: Any Mood vs. All					
7D – Caregiver Report	0 to 3.25	0.38	3.26+	3.26	.71, .78
7D – Self Report	0 to 2.17	0.49	2.18+	1.55	.74, .52
Step Two: BP vs. All					
Low Risk 7D					
7U – Caregiver Report	0 to 3.25	0.49	3.26+	2.11	.66, .68
7U – Self Report	0 to 2.67	0.73	2.68+	1.27	.64, .50
High Risk 7D					
7U – Caregiver Report	0 to 5.42	0.47	5.43+	1.94	.70, .64
7U – Self Report	0 to 5.30	0.51	5.31+	1.00	.71, .44

Table 5
Multinomial logistic regression with caregiver report

	Unipolar Diagnosis vs. Clinical Sample		*Bipolar Diagnosis vs. Clinical Sample*	
	B (SE)	Odds Ratio [95% CI]	*B* (SE)	Odds Ratio [95% CI]
Predictors				
Intercept	-4.78** (0.39)	-	-5.10** (0.39)	-
Sex	0.55** (0.16)	1.74 [1.26, 2.40]	0.46** (0.17)	1.58 [1.13, 2.21]
Race	0.05 (0.17)	1.05 [0.75, 1.47]	1.38** (0.17)	3.98 [2.85, 5.56]
Age	0.15** (0.02)	1.16 [1.11, 1.22]	0.03 (0.03)	1.03 [0.98, 1.08]
Number Axis I Diagnoses	0.66** (0.07)	1.94 [1.69, 2.24]	0.78** (0.07)	2.18 [1.89, 2.52]
7D – Caregiver Report	0.29** (0.02)	1.34 [1.28, 1.40]	0.24** (0.02)	1.27 [1.21, 1.32]
7U – Caregiver Report	-0.11** (0.02)	0.90 [0.86, 0.94]	0.12** (0.02)	1.14 [1.09, 1.19]

*$p<.05$, **$p<.01$

Table 6
Multinomial logistic regression with self report

	Unipolar Diagnosis vs. Clinical Sample		*Bipolar Diagnosis vs. Clinical Sample*	
	B (SE)	Odds Ratio [95% CI]	*B* (SE)	Odds Ratio [95% CI]
Predictors				
Intercept	-5.45** (0.80)	-	-7.10** (0.88)	-
Sex	0.38 (0.20)	1.46 [0.99, 2.17]	0.67** (0.22)	1.95 [1.26, 3.01]
Race	0.40* (0.20)	1.49 [1.01, 2.19]	1.32** (0.22)	3.73 [2.43, 5.71]
Age	0.25** (0.05)	1.28 [1.15, 1.42]	0.26** (0.06)	1.30 [1.16, 1.48]
Number Axis I Diagnoses	0.59** (0.09)	1.81 [1.53, 2.14]	0.74** (0.09)	2.09 [1.75, 2.50]
7D – Self Report	0.11** (0.02)	1.12 [1.07, 1.17]	0.06* (0.02)	1.06 [1.01, 1.11]
7U – Self Report	-0.04 (0.03)	0.96 [0.91, 1.01]	0.05 (0.03)	1.05 [1.00, 1.11]

*$p<.05$, **$p<.01$

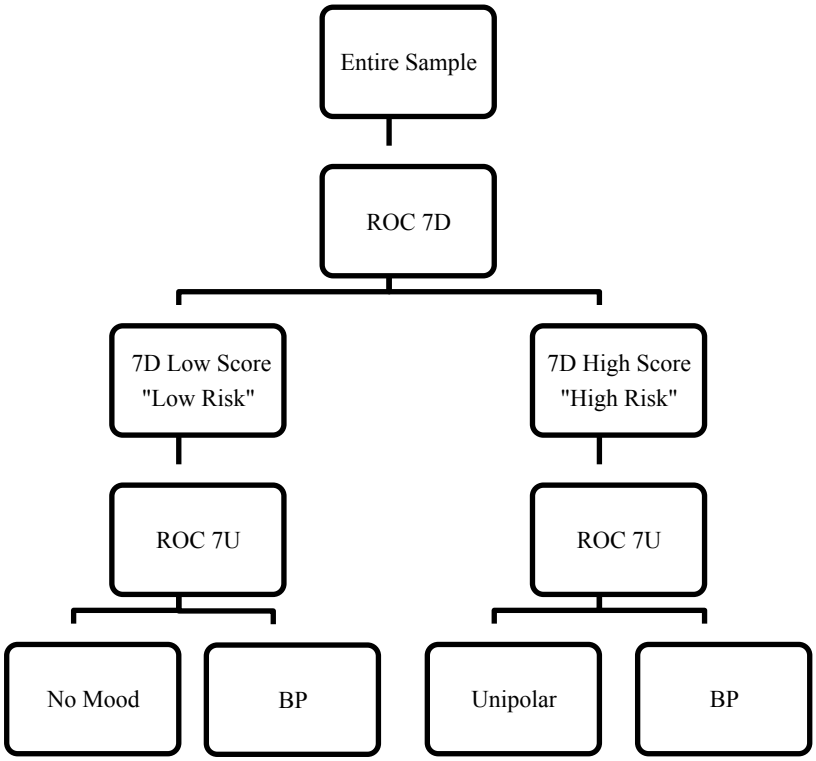

Figure 3. Proposed branching logic for two-step ROC analysis.

Current Episode \ Past episode	*Euthymia*	*Depression*	*Dysthymia*	*Cyclothymia*	*Hypomania*	*Mania*
Euthymia	No diagnosis	Depression	Dysthymia	Cyclothymia	BPNOS	BPI
Depression	Depression	Depression	Double depression	Unclear diagnosis	BPII	BPI
Dysthymia	Dysthymia	Depression, partial remission	Dysthymia	Cyclothymia	BPNOS	BPI
Cyclothymia	Cyclothymia	Unclear diagnosis	Cyclothymia	Cyclothymia	Cyclothymia	BPI
Hypomania	BPNOS	BPII	BPNOS	Cyclothymia	Bpnos	BPI
Mania	BPI	BPI	BPI	BPI	BPI	BPI

Figure 2. Mapping of combinations of current and past mood episodes to DSM-IV diagnoses of mood disorders.

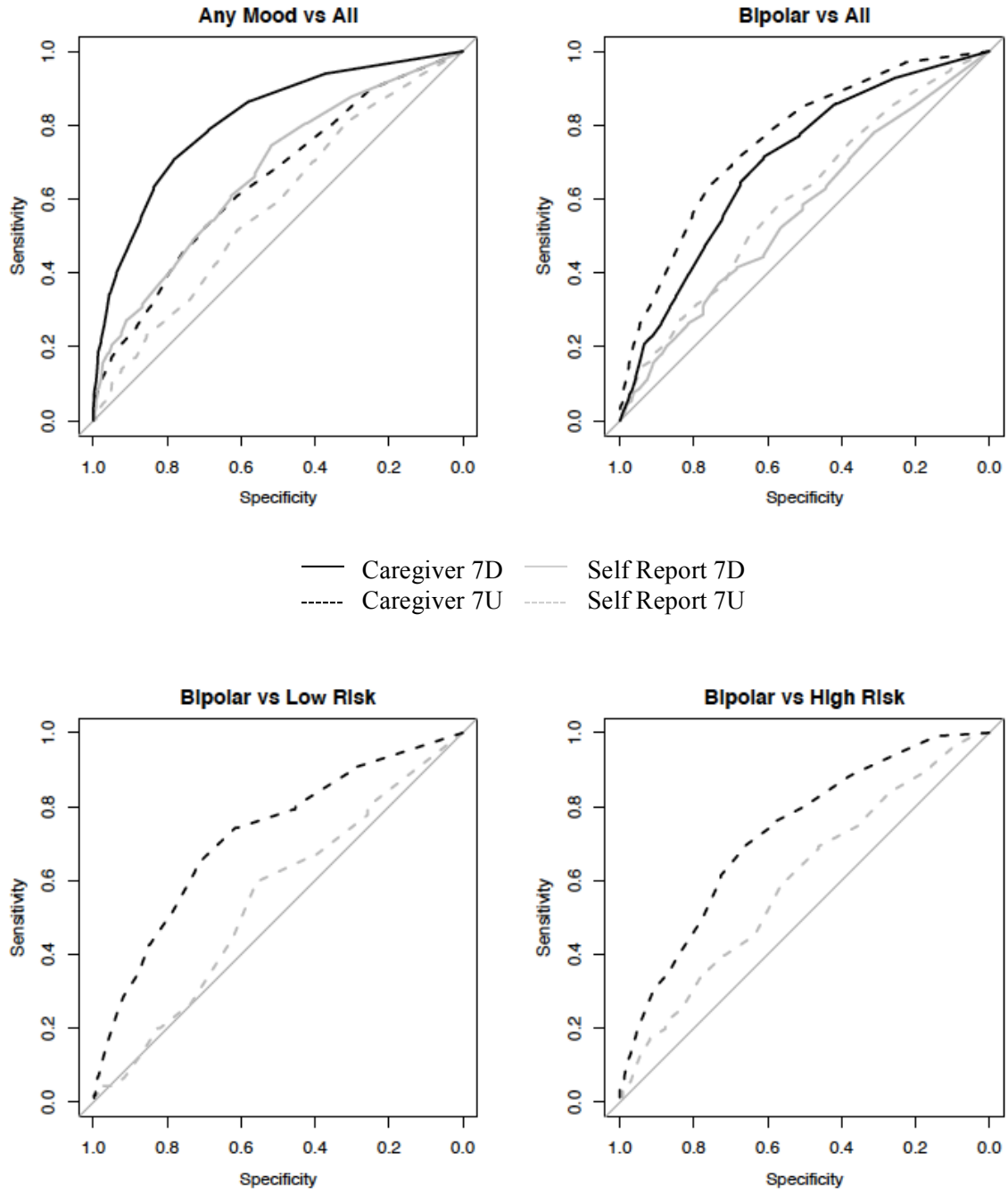

Figure 3
ROC Curves based on AUC estimates by informant report and target diagnosis.

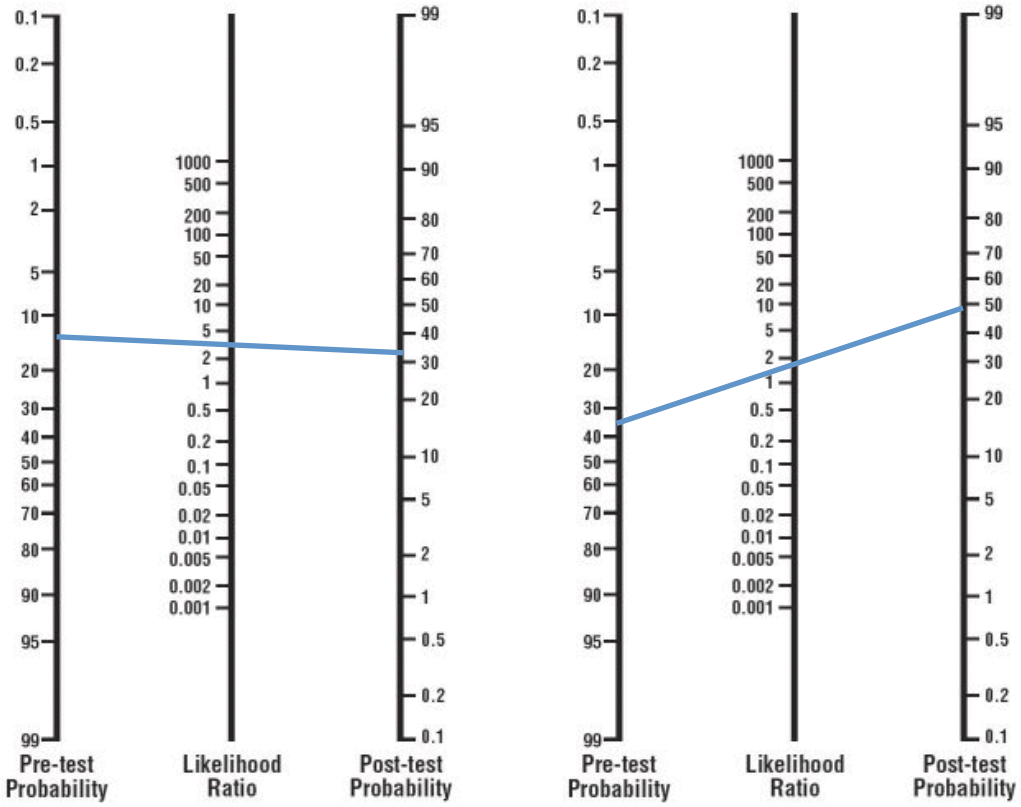

Figure 4. Nomogram example from clinical vignette.

REFERENCES

Akiskal, H. S., Bourgeois, M. L., Angst, J., Post, R., Moller, H. J., & Hirschfeld, R. (2000). Re-evaluting the prevalence of and diagnostic composition within the braod clinical spectrum of bipolar disorders. *Journal of Affective Disorders, 59*, s5-s30.

Alloy, L. B., Urosevic, S., Abramson, L. Y., Jager-Hyman, S., Nusslock, R., Whitehouse, W. G., & Hogan, M. (2012). Progression along the bipolar spectrum: A longitudinal study of predictors of conversion from bipolar spectrum conditions to bipolar I and II disorders. *Journal of Abnormal Psychology, 121*(1), 16-27. doi:10.1037/a0023973.

Altamura, A. C., Dell'Osso, B., Berlin, H. A., Buoli, M., Bassetti, R., & Mundo, E. (2010). Duration of untreated illness and suicide in bipolar disorder: a naturalistic study. *European Archives of Psychiatry and Clinical Neuroscience, 260*(5), 385-391. doi:10.1007/s00406-009-0085-2.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders : DSM-IV-TR*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM 5* (5th ed.). Washington, DC: American Psychiatric Association.

Angst, J., Gamma, A., Benazzi, F., Ajdacic, V., Eich, D., & Rossler, W. (2002). Toward a re-definition of subthreshold bipolarity: epidemiology and proposed criteria for bipolar-II, minor bipolar disorders and hypomania. *Journal of Affective Disorders, 73*, 133-146.

Arrasate, M., Gonzalez-Ortega, I., Alberich, S., Gutierrez, M., Martinez-Cengotitabengoa, M., Mosquera, F., . . . Gonzalez-Pinto, A. (2014). Affective dimensions as a diagnostic tool for bipolar disorder in first psychotic episodes. *European Psychiatry, 29*(7), 424-430. doi:10.1016/j.eurpsy.2013.07.005.

Axelson, D., Birmaher, B., Strobel, W., Goldstein, B. I., Ha, W., Gill, M. K., . . . Keller, M. (2011). Course of subthreshold bipolar disorder in you: Diagnostic progression from bipolar disorder not otherwise specified. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*(10), 1001 -1016.

Axelson, D., Birmaher, B., Strober, M., Gill, M. K., Valeri, S., Chiapetta, L., . . . Keller, M. (2006). Phenomology of children and adolescents with bipolar spectrum disorders. *Archives of General Psychiatry, 63*, 1139 - 1148.

Berk, M., Brnabic, A., Dodd, S., Kelin, K., Tohen, M., Malhi, G. S., . . . McGorry, P. D. (2011). Does stage of illness impact treatment response in bipolar disorder? Empirical treatment data and their implication for the staging model and early intervention. *Bipolar Disorders, 13*(1), 87-98. doi:10.1111/j.1399-5618.2011.00889.x

Berk, M., & Dodd, S. (2005). Bipolar II disorder: A review. *Bipolar Disorders, 7*, 11-21.

Berk, M., Dodd, S., Callaly, P., Berk, L., Fitzgerald, P., de Castella, A. R., . . . Kulkarni, J. (2007). History of illness prior to a diagnosis of bipolar disorder or schizoaffective disorder. *Journal of Affective Disorders, 103*(1-3), 181-186. doi:10.1016/j.jad.2007.01.027

Biederman, J. J. (1998). Resolved: Mania is mistaken for ADHD in prepubertal children. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*(10), 1091.

Bowden, C. L. (2001). Strategies to reduce misdiagnosis of bipolar depression. *Psychiatric Services, 52*(1), 51-55.

Bowden, C. L. (2005). A different depression: clinical distinctions between bipolar and unipolar depression. *Journal of Affective Disorders, 84*(2-3), 117-125. doi:10.1016/s0165-0327(03)00194-0

Bowring, M. A., & Kovacs, M. (1992). Difficulties in diagnosing manic disorders among children and adoelscents. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*(4), 611-614.

Danielson, C. K., Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003). Discriminative validity of the general behavior inventory using youth report. *Journal of Abnormal Child Psychology, 31*(1), 29-39. doi:10.1023/a:1021717231272

Depue, R. A., Krauss, S., Spoont, M. R., & Arbisi, P. (1989). General Behavior Inventory identification of unipolar and bipolar affective conditions in a nonclinical university population. *Journal of Abnormal Psychology, 98*(2), 117-126. doi:10.1037//0021-843x.98.2.117

Depue, R. A., Slater, J. F., Wolfstetterkausch, H., Klein, D., Goplerud, E., & Farr, D. (1981). A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: A conceptual framework and 5 validation studies. *Journal of Abnormal Psychology, 90*(5), 381-437. doi:10.1037/0021-843x.90.5.381

Findling, R., Youngstrom, E., Danielson, C. K., DelPorto-Bedoya, D., Papish-David, R., Townsend, L., & Calabrese, J. (2002). Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. *Bipolar Disorders, 4*, 34-42.

Findling, R. L., Gracious, B. L., McNamara, N. K., Youngstrom, E. A., Demeter, C. A., Branicky, L. A., & Calabrese, J. R. (2001). Rapid, continuous cycling and psychiatric co-morbidity in pediatric bipolar I disorder. *Bipolar Disorders, 3*(4), 202-210.

Findling, R. L., Youngstrom, E. A., Danielson, C. K., DelPorto-Bedoya, D., Papish-David, R., Townsend, L., & Calabrese, J. R. (2002). Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. *Bipolar Disorders, 4*(1), 34-42. doi:10.1034/j.1399-5618.2002.40102.x

Geller, B., Zimerman, B., Williams, M., Bolhofner, K., Craney, J., DelBello, M. P., & Soutullo, C. (2001). 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, 40*(4), 450-455. doi:10.1097/00004583-200104000-00014

Hillegers, M., Reichart, C., Wals, M., Verhulst, F. C., Ormel, J., & Nolen, W. A. (2005). Five-year prospective outcome of psychopathology in the adolescent offspring of bipolar parents. *Bipolar Disorders, 7*, 344-350.

Jenkins, M. M., Youngstrom, E. A., Washburn, J. J., & Youngstrom, J. K. (2011). Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology, Research and Practice, 42*(2), 121-129. doi:10.1037/a0022506

Judd, L. L., & Akiskal, H. S. (2003). The prevalence and disability of bipolar spectrum disorders in the US popoulation: re-analysis of the ECA database taking into account subthreshold cases. *Journal of Affective Disorders, 73*, 123-131.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., . . . Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(7), 980-988. doi:10.1097/00004583-199707000-00021

Keck, P. E., Kessler, R. C., & Ross, R. R. (2008). Clinical and economic effects of unrecognized or inadequately treated bipolar disorder. *Journal of Psychiatric Practice, 14*(2), 31-38.

Kim, E. Y., & Miklowitz, D. J. (2002). Childhood mania, attention deficit hyperactivity disorder and conduct disorder: a critical review of diagnostic dilemmas. *Bipolar Disorders, 4*, 215-225.

Mallon, J. C., Klein, D., Bornstein, R. F., & Slater, J. F. (1986). Discriminant validity of the General Behavior Inventory: An outpatient study. *Journal of Personality Assessment, 50*(4), 568-577.

Marchand, W. R. W. R. (2006). Delayed diagnosis of pediatric bipolar disorder in a community mental health setting. *Journal of Psychiatric Practice, 12*(2), 128.

McCombs, J. S., Ahn, J., Tencer, T., & Shi, L. (2007). The impact of unrecognized bipolar disorders among patients treated for depression with antidepressants in the fee-for-services California Medicaid (Medi-Cal) program: a 6-year retrospective analysis. *Journal of Affective Disorders, 97*(1-3), 171-179. doi:10.1016/j.jad.2006.06.018

McFadden, D. (1977). *Quantitative methods for analyzing travel behavior of individuals: some recent developments*: Institute of Transportation Studies, University of California.

Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R., Petukhova, M., & Kessler, R. C. (2007). Lifetime and 12-minth prevalence of bipolar spectrum disorder in the national comorbidity survey replication. *Archives of General Psychiatry, 64*(543 - 552).

Merikangas, K. R., He, J.-p., Burstein, M., Swanson, S. A., & Avenevoli, S. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(10), 980-989. doi:10.1016/j.jaac.2010.05.017

Mesman, E., Youngstrom, E. A., Juliana, N. K., Nolen, W. A., & Hillegers, M. H. (2017). Validation of the Seven Up Seven Down Inventory in bipolar offspring: screening and prediction of mood disorders. Findings from the Dutch Bipolar Offspring Study. *Journal of Affective Disorders, 207*, 95-101. doi:10.1016/j.jad.2016.09.024

Metz, C. E. (1978). Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine, 8*(4), 283 - 298.

Orvaschel, H. (1994). *Schedule for Affective Disorders and Schizophrenia for School-Age Children—Epidemiologic Version (K-SADS-E) 5.* Florida: Nova Southeastern University, Center for Psychological Studies.

Pendergast, L. L., Youngstrom, E. A., Merkitch, K. G., Moore, K. A., Black, C. L., Abramson, L. Y., & Alloy, L. B. (2014). Differentiating bipolar disorder from unipolar depression and ADHD: the utility of the general behavior inventory. *Psychological Assessment, 26*(1), 195-206. doi:10.1037/a0035138

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*(1), 77. doi:10.1186/1471-2105-12-77

Shrout, P. E., & Yager, T. J. (1989). Reliability and validity of screening scales: Effect of reducing scale length. *Journal of Clinical Epidemiology, 42*(1), 69-78.

Simon, G. E. (2003). Social and economic burden of mood disorders. *Biological Psychiatry, 54*(3), 208-215. doi:10.1016/s0006-3223(03)00420-7

Smith, D. J., Griffiths, E., Kelly, M., Hood, K., Craddock, N., & Simpson, S. A. (2011). Unrecognised bipolar disorder in primary care patients with depression. *British Journal of Psychiatry, 199*(1), 49-56. doi:10.1192/bjp.bp.110.083840

Spitzer, R. L. (1983). Psychiatric diagnosis: are clinicians still necessary? *Comprehensive Psychiatry, 24*(5), 399-411.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1-26. doi:10.1111/1529-1006.001

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological sicence can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1-26.

Van Meter, A. R., Moreira, A. L., & Youngstrom, E. A. (2011). Meta-analysis of epidemiologic studies of pediatric bipolar disorder. *Journal of Clinical Psychiatry, 72*(9), 1250-1256. doi:10.4088/JCP.10m06290

Venkatraman, E. S. (2000). A permutation test to compare receiver operating characteristic curves. *Biometrics, 56*. doi:10.1111/j.0006-341X.2000.01134.x

Wolkenstein, L., Bruchmuller, K., Schmid, P., & Meyer, T. D. (2011). Misdiagnosing bipolar disorder--do clinicians show heuristic biases? *Journal of Affective Disorders, 130*(3), 405-412. doi:10.1016/j.jad.2010.10.036

Youngstrom, E. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology, 39*(2), 204-221. doi:10.1093/jpepsy/jst062

Youngstrom, E., Arnold, E., & Frazier, T. (2010). Bipolar and ADHD comorbidity: Both artifact and outgrowth of shared mechanisms. *Clinical Psychology Science and Practice, 17*, 350-359.

Youngstrom, E., Meyers, O., Demeter, C., Youngstrom, J., Morello, L., Piiparinen, R., . . . Findling, R. L. (2005). Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disord, 7*(6), 507-517. doi:10.1111/j.1399-5618.2005.00269.x

Youngstrom, E. A., Arnold, L. E., & Frazier, T. W. (2010). Bipolar and ADHD comorbidity: both artifact and outgrowth of shared Mechanisms. *Clinical Psychology-Science and Practice, 17*(4), 350-359. doi:10.1111/j.1468-2850.2010.01226.x

Youngstrom, E. A., Birmaher, B., & Findling, R. L. (2008). Pediatric bipolar disorder: Validity, phenomenology and recommendations for diagnosis. *Bipolar Disorders, 10*, 13-13.

Youngstrom, E. A., Findling, R. L., Danielson, C. K., & Calabrese, J. R. (2001). Discriminative validity of parent report of hypomanic and depressive symptoms on the general behavior inventory. *Psychological Assessment, 13*(2), 267-276. doi:10.1037/1040-3590.13.2.267

Youngstrom, E. A., Frazier, T. W., Demeter, C., Calabrese, J. R., & Findling, R. L. (2008). Developing a 10-item mania scale from the parent general behavior inventory for children and adolescents. *Journal of Clinical Psychiatry, 69*(5), 831-839.

Youngstrom, E. A., Genzlinger, J. E., Egerton, G. A., & Van Meter, A. R. (2015). Multivariate meta-analysis of the discriminative validity of caregiver, youth, and teacher rating scales for pediatric bipolar disorder: Mother knows best about mania. *Archives of Scientific Psychology, 3*(1), 112-137. doi:10.1037/arc0000024

Youngstrom, E. A., Joseph, M. F., & Greene, J. (2008). Comparing the psychometric properties of multiple teacher report instruments as predictors of bipolar disorder in children and adolescents. *Journal of Clinical Psychology, 64*(4), 382-401. doi:10.1002/jclp.20462

Youngstrom, E. A., Murray, G., Johnson, S. L., & Findling, R. L. (2013). The 7 up 7 down inventory: a 14-item measure of manic and depressive tendencies carved from the General Behavior Inventory. *Psychological Assessment, 25*(4), 1377-1383. doi:10.1037/a0033975

Zweig, M. H., & Campbell, G. C. (1993). Receiver-operating Characteristic (ROC) Plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(4), 561-577.