



NIH PUBLIC ACCESS

Author Manuscript

Child Adolesc Psychiatr Clin N Am. Author manuscript; available in PMC 2010 August 4.

Published in final edited form as:

Child Adolesc Psychiatr Clin N Am. 2009 April ; 18(2): 353–ix. doi:10.1016/j.chc.2008.12.002.

The Assessment of Bipolar Disorder in Children and Adolescents

Eric A. Youngstrom, Andrew J. Freeman, and Melissa McKeown Jenkins

University of North Carolina at Chapel Hill

Abstract

The overarching goal of this review is to examine the current best evidence for assessing bipolar disorder in children and adolescents and provide a comprehensive, evidence-based approach to diagnosis. Evidence-based assessment strategies are organized around the “3 Ps” of clinical assessment: *Predict* important criteria or developmental trajectories, *Prescribe* a change in treatment choice, and inform *Process* of treating the youth and his/her family. The review characterizes bipolar disorder in youths - specifically addressing bipolar diagnoses and clinical subtypes; then provides an actuarial approach to assessment - using prevalence of disorder, risk factors, and questionnaires; discusses treatment thresholds; and identifies practical measures of process and outcomes. The clinical tools and risk factors selected for inclusion in this review represent the best empirical evidence in the literature. By the end of the review, clinicians will have a framework and set of clinically useful tools with which to effectively make evidence-based decisions regarding the diagnosis of bipolar disorder in children and adolescents.

There have been radical changes in our scientific understanding and clinical practices around the diagnosis of bipolar disorder in children and adolescents. Whereas the condition used to seldom be diagnosed before puberty, there has been a recent surge in rates of diagnosis such that a large proportion of psychiatrically hospitalized youths now carry clinical diagnoses of bipolar disorder¹, and there has been a more than 40 fold increase in rates of diagnoses over a 10-year period². There has been debate about whether the increase in diagnosis is primarily due to a correction of previous under-diagnosis, versus concerns that it is now overdiagnosed or even a case of “diseasemongering”³. Discussion has also focused on whether bipolar disorder in youth is the same illness as in adults, versus representing a different condition or perhaps a pediatric subtype⁴⁻⁶. Although the topic is still portrayed as controversial in the popular media, at this point more than 350 peer-reviewed publications have investigated different aspects of pediatric bipolar illness⁷. Growing evidence from clinical and epidemiological studies around the world indicates that bipolar disorder often first manifests in adolescence or earlier^{8,9}, and many apparent differences between adult and child presentations appear to be an artifact of definitional issues and not real variations in clinical presentation⁷. Prospective longitudinal studies also are documenting moderate to high levels of developmental continuity with adult bipolar disorder¹⁰⁻¹². All lines of evidence strongly indicate that bipolar symptoms in youths are associated with considerable impairment and warrant clinical attention.

The goal of the present review is to provide a step-by-step, evidence-based approach to the assessment of bipolar disorder in children and adolescents. The review is organized around clinical decision-making and then monitoring progress over the course of treatment. The

review does not discuss the potential merits of all the most commonly used instruments for psychological, educational, or psychiatric assessment; instead, it concentrates on those tools that have research supporting their validity with regard to pediatric bipolar disorder. The majority of the assessment tools routinely used for psychological¹³ and psychiatric evaluation have not been validated for work with pediatric bipolar disorder¹⁴. Rather than using standardized assessment batteries out of convention or habit, we believe that the assessment endeavor will perform best when each component is chosen based on its demonstrated validity and relevance to clinical intervention. Evaluation strategies should address one of the “3 Ps” of clinical assessment: 1) *Predict* important criteria or developmental trajectories, 2) *Prescribe* a change in treatment choice, or 3) Inform the *Process* of treating the patient or family¹⁵. Narrowing assessment batteries down in this manner has many benefits that include creating a strong link between assessment and treatment, reducing time and expense by eliminating unnecessary testing, and improving decisions and treatment outcomes by reducing “information clutter” and providing more focused information that directly pertains to the individual patient. The 3 Ps provide a rubric to help navigate the assessment process from establishing risk of pediatric bipolar disorder, confirming the diagnosis, informing treatment selection, measuring progress and outcome, and monitoring for relapse prevention.

Definitions: Mood Episodes and Bipolar Spectrum

The diagnostic criteria for mood disorders are unusual in that they require a two-stage evaluation^{16,17}. First, the clinician must evaluate the lifetime history of mood episodes, not just characterizing the current presenting problem. Only after gathering data about the possible occurrence of each type of potential mood episode over the lifetime can the clinician proceed to establishing the formal diagnosis. Diagnosing bipolar disorder requires this complexity, because the presentation of the illness can change dramatically as it transitions into different episodes.

Mood Episodes

The diagnostic mood episodes include major depressive episodes, dysthymic episodes, hypomanic symptoms, hypomanic episodes, manic episodes, and mixed episodes. These categories are not exhaustive in terms of phenomenology. Additional mood presentations are possible and frequently encountered in clinical practice, including mild depressions, mixed hypomanias, and periods of mood dysregulation that are too brief or mild to meet current criteria for an index episode¹⁸. However, the formal diagnosis of mood disorder is anchored to the index episodes, not the other clinical presentations. It is only after ascertaining both the present and past lifetime mood episodes that the clinician can diagnose mood disorders accurately. Table 1 shows how the combination of present and past episodes is often necessary to make a diagnosis on the bipolar spectrum. Unless the clinician inquires about past mood history, many cases of bipolar disorder will be misdiagnosed as unipolar depressive or dysthymic disorders – particularly given that people affected by bipolar disorders tend to spend more days depressed than manic, and are much more likely to seek services for depression than mania¹⁹. The situation may be somewhat different with pediatric bipolar disorder, both because referrals are more often initiated by the parent rather than the youth in outpatient settings, and because mania and mixed episodes appear to be more common in younger cohorts and then decrease steadily with age^{8,20}.

Diagnoses

Bipolar I, often considered the most serious form of bipolar illness, has received the greatest attention from the research community. As Table 1 makes evident, a bipolar I diagnosis only requires the presence of one manic or mixed lifetime episode¹⁷. Bipolar II disorder, in

contrast, requires two distinct mood episodes in order to assign the diagnosis: at least one major depressive episode and a hypomanic episode. Without systematic assessment for lifetime hypomanic episodes, bipolar II is very likely to be misdiagnosed as a unipolar depression²¹.

The DSM-IV includes cyclothymic disorder as another condition to be considered in the bipolar family of disorders. The diagnostic criteria for cyclothymic disorder require long periods of moderate mood disturbance. The depressive symptoms cannot become too severe, or else the diagnosis would change to a major depressive episode or bipolar II disorder. Similarly, the hypomanic symptoms cannot become too extreme; otherwise, if they meet criteria for a full manic episode, then the diagnosis would change to bipolar I. Cyclothymic disorder is difficult to distinguish from temperament; indeed, much research in the area has used rating scales assessing cyclothymic temperament²². The diagnosis of cyclothymic disorder is rarely used in clinical practice²³, nor is it tracked in most large epidemiological studies or clinical research samples^{8,10,24,25}. However, those studies that have investigated cyclothymic disorder have found that it is a highly impairing condition that warrants clinical attention²⁶⁻³³.

Bipolar “not otherwise specified” (NOS) is a fourth diagnostic option in the bipolar section of DSM-IV. Bipolar NOS is a residual category, intended to be used when bipolar features are present, but the clinical presentation does not fit into any of the three above categories. DSM-IV provides some examples of presentations that would be appropriate to code as bipolar NOS. These include having recurrent hypomanias without any lifetime history of manic, mixed, or major depressive episodes³⁴; having a disturbance in mood but with an insufficient number of the possible seven B criteria symptoms (e.g., elated mood plus one or two other symptoms; or irritable mood plus fewer than four other symptoms); or cases where the duration of the index mood episode is not long enough to satisfy the thresholds specified for hypomania (four days), or mania or mixed episodes (one week, or else severe enough to necessitate psychiatric hospitalization). Both the “insufficient number of symptoms”^{12,35,36} and the “insufficient duration” forms of bipolar NOS^{26,37} have been documented in multiple studies in both children and adults. Although the definitions do not identify identical sets of cases, the cumulative evidence shows that either definition is associated with considerable chronicity and clinical impairment. If the core feature of episodic mood disturbance is present, then most evidence suggests that bipolar NOS falls on the bipolar spectrum. In short, bipolar NOS (a) appears to be at least as prevalent as bipolar I in epidemiological and clinical samples, (b) has become well established as an impairing mood disorder, and (c) deserves clinical attention.

There is an important consideration about the potential overlap between bipolar NOS and cyclothymic disorder. In practice, most practitioners and researchers tend to lump cyclothymic disorder into the bipolar NOS category. Technically this is a departure from the official nosologies^{16,17}, and it adds to the heterogeneity that is found under the rubric of bipolar NOS. The combination of short durations for mood *states* combined with long lengths of *episode* should trigger careful evaluation of the possibility of a cyclothymic disorder.

When making any of the bipolar diagnoses, the clinician must rule out the possibility that the mood symptoms are due to schizophrenia, a general medical condition, or induced by a substance^{16,17}. The substance induced exclusion criteria creates the most challenges. Street drugs that have a strong dopaminergic effect can mimic the symptoms of mania, and hallucinogens can create symptoms that appear psychotic. A more subtle point is that manic symptoms secondary to the use of prescription medications, including antidepressant or stimulant medications, technically lead to diagnoses of “substance induced mania”. The

literature on psychotropic medications inducing mania is complex³⁸, but experts agree that manic symptoms emerging during the course of treatment always justify thorough evaluation of the possibility of bipolar diagnosis.

Additional Subtypes of Bipolar Disorder

In the pediatric literature, there has been much discussion about changes to criteria for youths or alternate definitions of bipolar subtypes⁶. Leibenluft and colleagues suggested the term “narrow phenotype” to indicate situations where the manic episode included symptoms of elated mood or grandiosity, consistent with the research operational definition of bipolar disorder used by Geller and colleagues^{6,39}. People often think that the term “narrow” connotes strict adherence to DSM criteria, but actually the “narrow” definition is more restrictive than DSM criteria (which would include hypomania or mania with predominantly irritable mood, so long as there were sufficient numbers of B-criteria symptoms co-occurring). In many samples, there is substantial overlap between the cases that would meet DSM criteria that would also satisfy the narrow definition^{37,40}. There is also considerably less research available based on the narrow criteria instead of the DSM criteria⁴¹. At the other extreme, the term “broad phenotype” has been used so widely and to refer to so many different things that it has become imprecise to the point of losing clinical utility. Because the evidence base is much stronger for DSM definitions than any alternate research definitions and DSM criteria guide clinical practice, the rest of this review will concentrate on DSM definitions (bipolar I, bipolar II, cyclothymic disorder, and bipolar NOS – clarifying whenever possible if the NOS specification is due to insufficient symptoms or insufficient duration).

Course Specifiers: Definitions and Clinical Relevance

Course specifiers have considerable clinical value in the context of bipolar disorder. Notations such as “bipolar II disorder, current episode depressed” provide important information about the nature of the illness and change some of the treatment options (i.e., prescription of different interventions for unipolar versus bipolar depression). There has been inconsistency about the use of the terms “cycling” and “rapid cycling.” These are often used to connote polarity switches. However, the DSM definition of rapid cycling denotes the occurrence of at least four or more distinct mood episodes (not changes in mood state within an episode) within the same year. Thus, rapid cycling might be thought of as “rapid recurrence” or “rapid relapse.” Indeed, when defined as four or more annual episodes, “rapid cycling” portends to a much more chronic course, higher rates of comorbidity and substance use, greater treatment refractoriness, and potentially greater risk of suicide. Thus the phenomenon of “rapid cycling/relapsing” satisfies both the predictive and prescriptive litmus tests for inclusion in an assessment of bipolar disorder. If the rapid switching between mood polarities that has been well-described in children is better construed as a mixed episode rather than multiple episodes; then the terms are being used consistently across the lifespan, and clinicians can better identify when there is a higher risk of relapse.

How Common Is Pediatric Bipolar Disorder? Baseline Risk and Prevalence

The first question that must be answered is, “How common is pediatric bipolar disorder, anyway?” At the time most practicing clinicians were trained, the conventional wisdom was that bipolar disorder affected only adults and perhaps some adolescents; and the vast majority of training programs still do not provide formal didactics about the assessment or treatment of pediatric bipolar disorder⁴². The prevalence of the disorder is an important starting point for clinical evaluation.

The traditional figure has been that bipolar disorder affects 1% of the adult population. This figure was often based on rates of bipolar I, and excluded all other DSM bipolar diagnoses. More recent epidemiological studies have found lifetime prevalences of bipolar I and II to be closer to 3 or 4%^{24,25}, and bipolar *spectrum* diagnoses appear to affect from 2.6% to 8.3% or more³⁶ of the general population (see Goodwin and Jamison for a review of twelve international studies)⁸. Unfortunately, epidemiological studies tend not to use strict DSM criteria for diagnoses, making it difficult to map findings directly onto clinical labels. Despite the varying definitions, it is clearly evident that (a) the bipolar spectrum is more common than generally thought, (b) the “soft spectrum” cases occur at least as frequently as does bipolar I in both community and clinical samples, and (c) the soft spectrum is associated with both immediate impairment and long term risk of poor outcomes on multiple measures^{10,36,37,43,44}.

A major caveat for the clinician is that epidemiological studies describe the incidence or prevalence of bipolar disorder in the general population. This is not the same thing as the frequency with which a practitioner will encounter bipolar disorder in clinical settings. Bipolar disorder is more common in outpatient settings than in nonreferred community samples, and bipolar is more frequent in settings providing more intensive services due to the acuity of the illness. Table 2 lists prevalence rates from multiple settings. The table also includes information about how the diagnoses were made, whether both parents and youths were interviewed, and other features that might influence the comparability of the estimates. Another limitation is that different groups and settings use somewhat different definitions of bipolar disorder, which also change the rates and their generalizability to other clinical settings. However, these rates still provide meaningful benchmarks against which clinicians can compare the rate of their bipolar disorder diagnoses. They also offer some indication of whether bipolar disorder is likely to be rare or common in a given setting.

Risk Factors

Research has identified multiple risk factors that might pertain to bipolar disorder. Tsuchiya and colleagues recently reviewed more than 100 studies evaluating more than 30 different risk factors. They concluded that only a family history of bipolar disorder is the only well-established risk factor for bipolar disorder that should warrant clinical attention⁴⁵. In studies of the offspring of bipolar parents, the risk of developing bipolar disorder appears to be at least five times higher than in the comparison groups⁴⁶, and estimates of the recurrence risk in adult samples indicate that the lifetime risk may be increased ten-fold⁴⁷. A recent review recommended that clinicians treat a history of bipolar illness in a first degree relative (biological mother, father, or full sibling) as increasing the risk of developing bipolar disorder by a factor of 5.0^{48,49}. Bipolar history in a grandparent, aunt, uncle, or half sibling would confer half as much risk (e.g., 2.5 times higher) based on the data suggesting that bipolar is a polygenic illness. These changes in likelihood are large enough to be informative in clinical assessment. At the same time, they are not so large as to make a diagnosis of bipolar disorder automatic; in fact, most people with an affected relative will not have bipolar disorder themselves. Table 3 lists other risk factors for pediatric bipolar disorder. These risk factors are less well established than family history, but are sufficient to prompt additional assessment of the possibility of a bipolar diagnosis.

There are several concerns that arise with regard to family history as a risk factor relevant to diagnosing pediatric bipolar disorder. These include: (a) the fact that the literature cannot yet disentangle genetic versus shared environmental familial factors, (b) the low diagnostic accuracy about bipolar diagnoses in general will undermine the sensitivity of family histories of bipolar disorder, and (c) bipolar disorder has historically been underdiagnosed in minority groups in the U.S.A., with it often misdiagnosed as schizophrenia or antisocial

behavior^{50,51}. For the purposes of formulating a diagnostic impression, it is not necessary to tease apart genetic versus environmental contributions. The poor sensitivity of family history means that failure to report a bipolar history cannot be assumed to be accurate, whereas positive reports of family history may be given greater credibility. The historical inaccuracy of bipolar diagnoses in minority groups means that clinicians need to inquire about mood symptoms whenever they hear about other mental health issues in relatives. Learning about prior treatment history of family members will also provide valuable information about attitudes towards treatment and adherence, and potentially about treatment response as well⁵².

Combining information: Impressionistic versus actuarial methods

How best can a clinician utilize information such as a positive family history of bipolar disorder, or test results? Clinical decision making is usually done based on expertise and impressionistic synthesis of different pieces of information about the individual patient. Case formulation and diagnosis are highly technical skills that integrate multiple variables and involve considerable amounts of training. Within a typical assessment framework, knowledge about the family history becomes one more piece of data to blend into the global diagnostic impressions, increasing concern about the likelihood of bipolar disorder, yet not guaranteeing the diagnosis. Family history is a “red flag,” ideally triggering other assessment procedures and helping build the case for a bipolar diagnosis when other confirming evidence emerges.

Using a Nomogram to Estimate Probabilities

It also is possible to use information about family history in a more quantitative manner. Evidence Based Medicine (EBM) advocates the use of Bayesian approaches for assessing the probability of a patient having a diagnosis⁵³. Bayesian methods focus on combining new information with the prior probability of a diagnosis in order to estimate a revised, posterior probability. Bayesian approaches have been available for centuries, but have not gained much popularity in clinical settings prior to the EBM movement. There are now a range of options for practitioners who want to use Bayesian methods. In addition to doing the computations by hand, there are also applets available on the Web or for personal digital assistants, and there are also “nomograms” that function like a probability slide rule, facilitating estimation of probabilities without requiring any computation, as shown in Figure 1.

Youngstrom and Duax⁴⁸ provide a detailed description of how to use a nomogram to estimate the probability of a youth having bipolar disorder when there is a family history of the illness. One first determines a starting probability, before considering the other information that will be synthesized with it. In the absence of any other information, the base rate of the diagnosis—as contained in Table 2--provides a helpful starting point⁵⁴. The base rate anchors the clinical decision with an objective consideration of whether bipolar disorder is going to be uncommon or fairly frequent in a clinical setting. The clinician locates the starting probability on the left-hand scale of the nomogram.

The middle line of the nomogram quantifies the impact of the new piece of assessment data, quantified as a “diagnostic likelihood ratio” (DLR)⁵³. Conceptually, the DLR indexes the change in risk of a condition by comparing the rate at which the assessment event (such as a positive family history, or a high test score) occurs in cases with bipolar disorder to the rate of occurrence for the same assessment event in cases without bipolar disorder. In other words, the DLR is the ratio of the *sensitivity* of the assessment to bipolar disorder (out of 100 cases with bipolar disorder, how many would obtain a positive assessment result), divided by the *false alarm rate* (out of 100 cases that do not have bipolar disorder, how

many would also “falsely” obtain a positive assessment results – the opposite of the tool’s *specificity* to the diagnosis). The DLR is the change in the odds of having a bipolar diagnosis. The nomogram avoids needing to perform calculations to combine the starting probability with the DLR. Instead, the clinician finds the DLR value on the middle column of the nomogram, and then connects the dots between the first line (the starting probability) and the second line (the DLR) and extends the line across to the third, right-hand scale of the nomogram, which provides the revised probability estimate.

For example, any clinician evaluating a youth coming to an outpatient clinic whose mother has been diagnosed with bipolar II could use the nomogram in the following manner. First, the clinician would select the base rate of bipolar disorder, either using local historical information about the rate of diagnosis at his/her clinic, or by finding a published estimate from a similar setting. The estimates listed in Table 2 suggested the base rate of 6% for bipolar spectrum disorders in outpatient clinics. Thus the clinician would put a dot at 6% on the left-hand line of the nomogram. Diagnosis of a bipolar disorder and a first degree relative increases the risk of a bipolar diagnosis in the youth by a factor of five to 10. The clinician opts to use the more conservative estimate, and marks the five on the middle line of the nomogram. Connecting the two dots and extending the line across the right-hand side of the nomogram yields a probability estimate in the vicinity of 24%, indicating that the youth has approximately a one in four chance of having bipolar disorder. Alternately, this value can be interpreted as meaning that roughly 24 out of 100 youths presenting to an outpatient clinic with a family history of bipolar disorder will themselves meet criteria for a bipolar spectrum diagnosis. If the clinician had picked the more liberal estimate of a tenfold increase due to family history, then the resulting risk estimate would have been roughly 39%. Comparing these two estimates illustrates several advantages of using the nomogram (or other Bayesian methods): (1) combining probabilities and risk factors is not an intuitive or linear process; (2) it is easy for clinicians to play “what if” scenarios by changing their starting assumptions or their choice of weight to assign to risk factors --referred to as “sensitivity analysis” in the EBM literature⁵³; and (3) the results from the nomogram fall along a continuum and communicate more accurate information about the degree of diagnostic certainty. One of the major pitfalls of diagnostic testing is that results are prone to misinterpretation, especially when test findings are treated as black and white statements about the patient’s status. The nomogram approach keeps the shades of gray. In this example, a “black and white” approach to testing would either focus on the test positive result (family history), or on the posterior probability being below 50%. Focusing only on the family history, or treating it as if it were synonymous with a bipolar diagnosis in the child, would be inaccurate in more than 3 out of 4 cases. The alternative would be to conclude that family history is not diagnostically useful in outpatient settings, because even when present, most youths will still not have bipolar disorder. Even when conducting a sensitivity analysis using two different estimates of risk, the results are consistent in showing that this particular combination of factors put the youth at moderate risk (24-39%) of having a bipolar spectrum illness. These numbers quantify the earlier statement that positive family history is a “red flag” that should initiate more comprehensive evaluation of a possible mood disorder.

How Does the Accuracy of Impressionistic Versus Statistical Methods Compare?

The actuarial/statistical approach to interpreting assessment information is unfamiliar to most clinicians, and also contrasts sharply with more intuitive approaches to interpretation. However, the literature is unambiguous that simple statistical approaches, such as the nomogram method, consistently outperform typical clinical judgment^{54,55}. The superiority of statistical approaches has been demonstrated more than 130 times, in disciplines spanning economics and education as well as clinical decision-making⁵⁶. Cognitive science is

beginning to elucidate reasons why even simple statistical approaches perform better. The culprits are “heuristics,” cognitive shortcuts that facilitate the rapid identification and interpretation of information^{57,58}. These heuristics help the brain process large volumes of complex information swiftly, but they also lead to systematic and predictable biases. The human brain pays attention to cues of risk, for example, and to err on the side of high sensitivity at the expense of false alarms. Though highly adaptive in situations where a failure to detect risk could result in death, the high sensitivity to risk may lead to overestimates of rare but risky events in clinical settings⁵⁹. A variety of other heuristics beset clinical judgment, including availability heuristics (such as noticing more bipolar symptoms in patients after repeatedly hearing about the rise in diagnosis in the popular press).

Many of these heuristics are likely to be relevant to the clinical diagnosis of bipolar disorder, suggesting that typical decision-making would be vulnerable to at least as much bias and error as described in the larger decision-making literature. In fact, emerging evidence indicates that clinical diagnoses often have low accuracy with regard to bipolar disorder, including long delays between the emergence and recognition of symptoms⁶⁰⁻⁶², cyclical trends where the diagnosis goes in and out of “fashion” compared to schizophrenia⁶³, low agreement with systematic research diagnoses of bipolar disorders⁶⁴, and large regional differences in the tendency to diagnose mania or in ratings of severity⁶⁵⁻⁶⁷. Coding videotaped interviews revealed a strong tendency for American clinicians to rate manic symptoms as more severe than British or Asian Indian clinicians⁶⁶, and ratings of clinical vignettes showed that American psychiatrists were more likely to classify ambiguous clinical presentations as “bipolar” versus the rates identified by British clinicians⁶⁵. More encouragingly, another vignette study has found that people can learn the nomogram approach quickly, and that applying the nomogram to the same clinical vignette results in significantly more accurate estimates of bipolar risk, greater consistency and agreement about the degree of risk (i.e., much smaller range of opinion, and smaller standard deviations), and a marked reduction in overdiagnosis of bipolar disorder⁶⁷. Similar improvements in decision-making have been documented in numerous other areas of medicine⁵³.

Questionnaires and Checklists

Questionnaires and behavior checklists are an important tool in the kit of pediatric healthcare professionals. They offer an inexpensive, systematic way of gathering information, potentially from multiple sources (e.g., teachers as well as parents or youths). The instruments can cover a broad range of areas of functioning and impairment, or they can drill deeper into more narrowly defined areas, helping to clarify diagnosis or establish the severity of problems. No single instrument will be equally suited to all of these diverse applications. What follows is a brief overview of the evidence pertaining to questionnaires and checklists with regard to pediatric bipolar disorders.

Broadband Checklists

“Broadband” checklists cover a wide range of behavior problems. Both empirically derived versions (e.g., the Achenbach System of Empirically Based Assessment, including the Child Behavior Checklist)⁶⁸ and DSM-oriented versions (e.g., CSI)⁶⁹ include subscales dealing with aggressive behavior, depression, anxiety, attention problems, social problems, and thought disorder or psychotic symptoms. The empirically derived versions also include superordinate scales that measure more global levels of externalizing and internalizing problems. Some versions provide scoring algorithms to map onto potential DSM diagnoses⁶⁹, and others include age and sex-based norms, comparing the level of behavior problems to typical levels of functioning for peers. Few of these instruments include a mania scale,

reflecting the historical fact that most item pools were generated before there was concern about the possibility of pediatric bipolar disorder. The exceptions still tend to only include the mania items in the adolescent version⁶⁹ or only on the self-report (and not parent or teacher report) versions (BASC)⁷⁰.

There are at least three major roles that a broad-band instrument can play in the context of evaluating PBD: (1) high externalizing scores can trigger further assessment substituting other procedures for evaluating mania; (2) low scores can substantially reduce the probability that a case has PBD; and (3) broadband measures provide an inexpensive method of gauging the range of associated problems and comorbidities frequently seen with PBD. The CBCL is the most thoroughly investigated measure with regard to PBD, and evidence consistently shows that youths with PBD show elevated scores on multiple scales, including the Externalizing Problems broadband score^{71,72}. However, in spite of PBD elevating average scores on several scales, from a diagnostic perspective it is the Externalizing Problems that convey the most information. After controlling for Externalizing scores, no other scale or combination of scales provides incremental validity⁷³. Most cases with PBD will show high Externalizing scores (i.e., they are sensitive to PBD), but high scores are also associated with many other conditions (i.e., they are not specific to PBD). This sets up an asymmetry, where low Externalizing scores are often decisive at ruling bipolar disorder out, but high scores are ambiguous⁷⁴. Because of this, high scores should be treated as another warning sign, leading to deeper investigation of potential bipolar disorder. On the other hand, low scores will often decrease the risk enough to effectively rule bipolar disorder out, unless there are several countervailing risk factors and clinical signs. Table 4 provides the DLRs associated with low, moderate, and high scores on the CBCL as well as the Achenbach TRF and YSR. More diagnostic information can be wrung from tests by estimating DLRs for multiple segments corresponding to low, medium, and high scores (as opposed to the common practice of setting a single threshold)⁵³. The low scores on the CBCL are more powerful at reducing probability of bipolar (DLR = .04) than extremely high scores are at increasing risk (DLR = 4, versus a 25 for a low score *reducing* risk)⁵³. Finally, regardless of whether the behavior problems represent true comorbid diagnoses, versus elements of a “core phenotype” of pediatric bipolar disorder or secondary consequences of the illness, the other clinical syndrome scales on broadband measures provide valuable information about functioning and other potential targets for treatment. For example, severe attention problems and chronic hyperactivity often require adjunctive treatment with stimulants even after mood stabilization has occurred⁷⁵⁻⁷⁷; and the social problems associated with PBD also respond well to targeted interventions^{78,79}.

Mania specific measures

Measures of mania for youths have proliferated over the last decade (see Table 4). They vary widely in terms of item content, reading level, and degree of validation. Although many are brief and most are in the public domain, the rarity of PBD and the false positive rates produced by all of the tests preclude recommending the use of any mania checklist as a core component of outpatient assessment batteries. However, the best available measures are markedly more specific to PBD than the broadband instruments, suggesting a cost-effective, two stage approach to assessment (see Figure 2)⁴⁹. First, the clinician would gather general developmental history and family history, which would include an assessment of several risk factors for PBD. Second, they would give a broadband measure as a way of getting a rapid scouting report about a wide array of clinical domains. If the family history and externalizing scores both were low risk, then bipolar disorder is effectively ruled out (Figure 2). If either the family history is significant for bipolar disorder or the Externalizing score is high, then the clinician would supplement the assessment battery with a mania-specific measure. At present, the best validated and most discriminating instruments are the PGBI³³

and its 10 item mania form⁸⁰, the Parent MDQ^{81,82}, and the CMRS⁸³ and its 10 item form⁸⁴. Other instruments have either not performed as well, or they have not been validated under similarly generalizable clinical circumstances⁸⁵. These three instruments produce functionally interchangeable results in terms of diagnostic assessment. As new papers are published, test users should compare the instruments not only on their area under the curve (AUC) in receiver operating characteristic (ROC) analyses (which combines the diagnostic sensitivity and specificity into a single summary score), but also on the quality of the study, sample, and reporting⁸⁶.

Interpreting Test Scores

The clinician can choose to interpret test scores from rating scales in a number of ways. The most common method is a categorical, impressionistic interpretation, where scores are grouped into “high” and “low” ranges. It is also possible to be more formal about the quantification of risk information conveyed by test results, using the same array of options as described earlier when interpreting family history. Current thinking in EBM is that the use of diagnostic likelihood ratios is a preferred strategy⁵³. The DLR for a test result is the percentage of cases with bipolar disorder divided by the percentage of nonbipolar cases scoring in the same range. If a publication provides the sensitivity and specificity, then it is straightforward to convert these values into a pair of DLRs for scores above and below the threshold⁵³. Table 4 includes the DLRs for all available tests with regard to PBD at the time of writing.

Perusal of the DLR values leads to several observations. Many of the DLRs associated with low test scores are values lower than 1.0. A DLR of 1.0 indicates that the test result or risk factor did not change the probability of a bipolar diagnosis, because the score is equally likely to occur in both bipolar and nonbipolar reference groups. DLRs smaller than 1.0 reflect that the score is much more likely to occur in nonbipolar cases, thus reducing the likelihood that the current client has bipolar disorder. Whereas a value of 2.0 would signify a doubling of the odds of a bipolar diagnosis, a value of 0.5 would convey a similar change in odds in the opposite direction. DLRs greater than 10 or smaller than 0.1 are often decisive pieces of information⁵³: They can change a prior probability of 50% (even odds) to more than 90% or less than 10% posterior probability. These benchmarks lead to the additional observation that available instruments are more powerful at decreasing the likelihood of bipolar disorder than at increasing it (i.e., there are many test results yielding DLRs less than 0.1, but few with DLRs greater than 10, and none that have been validated in samples with a high degree of clinical generalizability). Small DLRs can actually play a valuable role by reducing the tendency to overdiagnose PBD^{1,2,67}.

To use these DLRs with a nomogram, one follows the same procedure as described with the interpretation of family history, only using the test score’s DLR as the estimate on the middle line of the nomogram⁸⁷. When multiple DLRs are available, such as when both family history information and a CBCL Externalizing score are available, then all pieces of information can be combined within the nomogram framework. The sequence does not matter. Family history could be considered first, or the test score; or the DLRs could be multiplied together and the product used instead during a single pass through the nomogram. Algebraically, these are all equivalent scenarios. This degree of flexibility is extremely valuable clinically, though, as often it will not be possible to obtain some pieces of assessment information for a specific case, and the order with which clinical data become available often varies across cases. In contrast, other actuarial methods such as lookup tables, logistic regression, or decision trees require that all the component variables be measured for each case, and that they be applied in a specified combination or sequence. Building on the earlier example, if a very high CBCL Externalizing score ($T=83$) was added

to the case with a positive family history of bipolar disorder, then the current probability (24% for based on a first degree relative increasing the risk of a bipolar diagnosis by a factor of 5) would be entered on the lefthand line of the nomogram. Table 4 provides the DLR associated with the *T*-score (DLR = 4.3 for an adolescent scoring this high on the CBCL). Combining the prior probability and the DLR algebraically yields an estimate of 58%. Using the nomogram adds some imprecision, both because reference points need to be visually interpolated, and because of error connecting the dots; but clinical estimates using the nomogram still wind up being centered around the best estimate and are dramatically more precise than when clinicians interpret the same information impressionistically⁶⁷.

Either a high score on a broadband measure or a positive family history of bipolar disorder would justify the addition of one of the more specific mania scales to the assessment process. However, when the same person fills out two rating scales, only one of them should be incorporated into the formal assessment process, whether it be the nomogram or another method of combining risks¹⁴. The scores on the questionnaires will be highly correlated with each other by virtue of coming from the same source, and thus will yield redundant information. Treating multiple questionnaire scores as if each was introducing new information will create bias in the probability estimates. The bias can be substantial when the scores are highly correlated ($r > .5$), which will often be the case when the same person fills out multiple instruments, even when they measure different constructs⁸⁸. As a result, the clinician should take the most valid piece of information available from the informant and substitute it into the nomogram cycle, ignoring other scores from the same source¹⁴. In our case example, the high CBCL score would cause the clinician to ask the caregiver to complete the Parent GBI. A very high score on this tool (e.g., a raw score of 51; see Table 4) has a DLR of 9.2. This would replace the DLR of 4.3 from the CBCL completed by the same caregiver. Combining the DLR of 9.2 with a prior probability of 24% (based on the family history and the base rate of PBD in outpatient settings) generates an estimate of 58% risk.

Permuting the possible combinations of DLRs from the combination of family histories and test scores yields a range of 8 to 24 probability estimates per test. When also accounting for the differences in base rate across clinical settings, the number of distinct probabilities will exceed 100 for each test x family history x setting combination. This reveals another advantage of the nomogram approach compared to generating tables of estimates for each configuration, or versus trying to weight the information sources intuitively. The tandem of family history and rating scales is powerful enough to move the probability estimate of a PBD diagnosis to less than 1% (no family history plus a low score on a broadband) or as high as 85% (using a more aggressive estimate of 10-fold increase in risk due to family history, plus a very high score on a PGBI or comparable tool). Thus, an evidence-based approach to assessment can rule PBD out in many cases, and reduce the tendency to overdiagnose bipolar disorder; but even high risk combinations are not sufficiently accurate for diagnosis of bipolar disorder to replace a careful symptom level assessment of bipolar disorder.

Cross Informant Issues

There has been a substantial amount of research on the validity of youth self report and teacher report as well as caregiver report (almost always mothers) about bipolar disorder. Findings consistently show the greatest validity for parent report, which shows significantly larger effect sizes than youth or teacher report in all published studies where the same instrument is available from multiple informants^{73,82,89,90}. Examining Table 4 reveals that the DLRs for parent report are consistently larger than the DLRs for youth or teacher report on the same instruments. The greater validity of parent report persists even when the parent has a diagnosed mood disorder⁹¹.

That parent report outperforms self-report contradicts conventional wisdom that self-report is a better source of information about mood disorders⁹². The lower validity of youth report appears to be due to a combination of mania compromising insight into one's own behaviors⁹³ and manic symptoms tending to disturb others before the affected individual perceives them as problematic⁹⁴. The low validity of teacher report persists even when teachers complete mania-specific rating scales⁹⁵. Of note, the agreement between parents and youths or parents and teachers is actually significantly higher than typical for cross-informant agreement^{91,96}, and youths and teachers report significantly more behavior problems in PBD cases than would be predicted based on the parent's level of concern alone⁹¹. The challenge is the difficulty of intuitively appreciating what a cross-situational correlation of .2 or .3 might look like at a case level, so instances where dyadic agreement is actually good are often misinterpreted as one person having exaggerated concerns. Clinicians will often encounter families where the parent reports more mood issues than the teacher or youth. The evidence-based approach to these discrepancies is not to automatically discount the parent report, but rather to systematically gather additional information to evaluate the possibility of PBD¹⁴. At a statistical level, youth and teacher report provide only modest—and often insignificant—incremental validity after controlling for parent report^{73,94}. However, the correlation between parents and youths or teachers is sufficiently low that these can be treated as functionally separate sources of information and combined within the nomogram framework. Evidence also suggests that cases where mood symptoms are noticeable across informants and settings may have greater impairment⁹⁷, also justifying the effort required to collect multiple perspectives. Teacher report can add useful information about the degree of problems in the school setting⁹⁵, and youth report adds data about the degree of insight into problems and motivation for treatment, both providing helpful prescriptive information to guide intervention.

Deciding to Intervene: Crossing the Treatment Threshold

The clinical process has three different options with regard to diagnosis: Ruled out, sufficiently well-established that treatment for the condition should begin, or else possible but not established. In EBM, the probability of diagnosis theoretically ranges from 0 to 100%, and there are two thresholds that separate the three different clinical options (see Figure 2)⁵³. The Test-Wait threshold separates the low risk zone - diagnosis is effectively ruled out - from the indeterminate middle range. The Test-Treat threshold demarcates the zone where probability is high enough to initiate treatment. Figure 2 does not specify probability levels for these thresholds. In practice, the location of the threshold should take into consideration the risks and benefits associated with treatment, as well as patient preferences. There is a formal framework for collecting and incorporating these utilities into adjusted thresholds⁹⁸, with perhaps the easiest approach described in an EBM handbook⁵³. As diagrammed in Figure 2, the combination of family history and questionnaire data will be sufficient to rule bipolar disorder out when both indicate low risk. High risk cases, with positive family history and high scores on mania measures, will still fall below the Test-Treat threshold – especially when considering the stigma, treatment burden, and potential side effects attendant to recommended treatments for PBD⁹⁹. Within an EBM framework, probabilities falling between the Treat and Wait thresholds indicate the need for continued assessment. New assessment data then gets combined with the current probability until the revised probability crosses the Wait threshold (ruling the diagnosis out) or the Treat threshold (ruling the diagnosis in).

Doses of Treatment Model

Psychology has long had a model of “levels of intervention,” where primary preventions might be offered to everyone in order to avoid onset of an illness, secondary interventions might be offered only to targeted high-risk groups, and tertiary interventions would

deployed for cases already manifesting a disorder¹⁰⁰. Preventive measures need to be low-risk and low cost if they are going to be applied widely regardless of risk. Tertiary interventions can be higher risk and expense because they have been reserved for cases with established diagnoses. This “levels of intervention” model can be mapped onto the EBM diagnostic threshold model¹⁵, as shown in Figure 2, right-hand side. Synthesizing these two models creates a set of assessment and treatment recommendations for each of the three ranges of probability for a bipolar diagnosis. Instead of labeling cases having mid-range probabilities as “indeterminate,” they can be called “moderate risk” and treatment using low-risk methods can start. Techniques such as psychotherapy, dietary supplementation, and improved sleep hygiene all might be tried for cases in this range. So long as the burden, risks, and costs are low, then treatments that are nonspecific or potentially preventative can be used, even while assessment continues and clarifies diagnostic impressions.

Leaving Limbo: Ruling Bipolar Disorder In or Out

Even when multiple risk factors are present, a clinician cannot assume that a PBD diagnosis has been established. How best should they proceed? The following suggestions provide an overview of strategies that help to confirm the diagnosis in high risk cases, or to rule a bipolar diagnosis in or out for cases that fall in the intermediate range of risk.

Diagnostic Interviews

Diagnostic interviews remain the standard of practice for determining clinical diagnosis. The typical unstructured diagnostic interview is prone to a variety of heuristics that render its reliability quite low¹⁰¹. For PBD, circumstances are likely to worsen the already typically poor degree of inter-rater, due to issues such as the lack of formal training in recognition of PBD, the usage of different operational definitions, and the controversy around the diagnosis. Structured diagnostic interviews avoid some of the shortcomings of informal interviews, including systematic coverage of relevant symptoms and formal algorithms to make DSM diagnoses¹⁰¹. However, most structured interviews were designed before PBD was considered a serious possibility in youths, with the consequence that many pediatric structured interviews do not include a mania module, and those that do might include few if any modifications to the probes or anchors to facilitate recognition in pediatric cases¹⁴.

For this reason, semi-structured diagnostic interviews such as the KSADS have become the accepted standard for PBD research¹⁰². Different versions of the KSADS have demonstrated good reliability and validity with regard to PBD^{103,104}. There are some obstacles hindering the widespread clinical adoption of semi-structured interviews. These include the necessity for extensive training, or else the semi-structured aspect opens the door for differences in clinical judgment to undermine the reliability⁶⁶, as well as the substantial amount of time required to administer and score the interview. A full KSADS can take anywhere from 2 to 8 hours to complete with a family, with administrations by experienced clinicians often averaging around 3 hours for typical cases. However, more streamlined versions of semi-structured interviews deserve serious consideration as a potential component of clinical assessment¹⁰⁵. The emerging literature around clinical diagnosis of PBD suggests that there may be great need for semi-structured approaches despite the increased expense and burden involved. Medicaid and other providers will often reimburse for the diagnostic assessment time if medical necessity has been demonstrated. The framework described here – starting with a combination of rating scales and family history – provides strong documentation of medical necessity for such additional evaluation. Clinicians who want to adopt semi-structured interviews as part of their assessment portfolio should pick an instrument that covers manic and depressive symptoms thoroughly, includes developmentally appropriate anchors, and supports the diagnosis of “spectrum” conditions

(such as bipolar II, cyclothymic disorder, and bipolar NOS), as these will be more common than bipolar I, yet impairing enough to represent major clinical concerns.

“Handle” Symptoms

Not all symptoms carry equal weight towards making a diagnosis of mania. The DSM-IV and ICD criteria give greater emphasis to elated mood instead of irritable mood. Elated mood requires only three additional symptoms to support a diagnosis of mania, versus four additional symptoms for irritable mood¹⁷. This policy acknowledges that elated mood has greater diagnostic specificity to mania, whereas irritable mood is diagnostically nonspecific. Research suggests that decreased need for sleep, unstable self esteem and grandiosity, hypersexuality, racing thoughts, and psychotic symptoms are all relatively specific to PBD⁷. The sensitivity of each of these symptoms is low enough that none should be required for making a diagnosis of bipolar disorder, or else somewhere between a quarter and two thirds of bipolar cases might be excluded⁷. Each of these symptoms is also liable to occur in at least one other condition likely to be encountered in many clinical settings. For example, hypersexuality can be a sign of sexual abuse, and inflated self-esteem is frequently seen in conduct disorder⁷. However, a clinician can learn how these symptoms often manifest in the context of PBD¹⁰⁶, and careful probing around these symptoms is an important component of refining diagnostic impressions. Evidence that any of the possible manic symptoms occur episodically, as opposed to chronically, or that they fluctuate with changes in mood and energy, heighten the suspicion that they are due to a mood disorder rather than a more chronic condition such as ADHD¹⁰⁷. If the symptom occurs with an unusual Frequency, if the Intensity is excessive, if the Number of occurrences within an episode is extreme, or if the Duration is exceptional compared to age appropriate behavior, that also helps build the case in favor of a mood diagnosis^{99,108}.

Extending the Window of Assessment

Another crucial strategy to improve the detection of PBD is to extend the window of assessment beyond the conventional single session of intake assessment⁴⁹. Relying on a single panel of information focused on the presenting problem will rarely be enough to allow a firm diagnosis of PBD.

Developmental History

Gathering a developmental history is a routine component of pediatric assessment. Its role is especially helpful when evaluating the possibility of PBD. In addition to gathering data about the family psychiatric history, pre- and perinatal risk factors and complications and temperamental characteristics all deserve consideration¹⁰⁸. Developmental trajectories can help distinguish between chronic conditions such as ADHD (which is formally required to have onset before age seven) versus episodic mood presentations¹⁰⁷. Even though not all authorities concur that PBD will always have an episodic presentation, identifying episodic presentations still carries treatment utility by suggesting different intervention strategies^{52,109}. The most intensive form of retrospective information gathering would be to complete a retrospective life chart¹¹⁰. The retrospective life chart is a tool that asks the family to reconstruct a week-by-week summary of the youth's past mood and energy levels, using a variety of anchors and techniques to facilitate accurate recall. The retrospective life chart can yield valuable information about the chronicity versus episodicity of mood presentation, and it can help to identify triggering events. However, the time and effort involved is substantial, and retrospective memory is subject to several sources of bias. Clinically, the costs and possible benefits need to be balanced on an individual case basis before adding a prospective life chart.

Extending the Assessment Window Forward

The other way of moving beyond a single-session intake is by extending the window of assessment forward in time. There are many means of accomplishing this. They include starting with a diagnosis of “rule out bipolar disorder” based on the initial intake, and following up with additional assessment to clarify the diagnosis. The EBM threshold model operationalizes this concept by indicating continued assessment for as long as the probability of PBD falls between the Wait and Treat thresholds⁵³.

Another approach is to shift to a “dental model” of assessment, where ongoing “checkups” are scheduled to gauge mood and energy over the course of treatment⁴⁹. At a minimum, these could consist of asking the patient at each visit about their mood and energy since the last visit. Alternately, the family could complete brief rating scales every few weeks over the course of treatment^{76,77}. The clinician could also quantify impressions using ratings such as the Children’s Global Assessment Scale (CGAS)¹¹¹. At the most intensive, the clinician could suggest doing prospective life charting¹¹⁰. In a prospective format, the patient records changes in mood and energy on a daily basis, and also notes any coinciding events. There are free prospective life charts for use with youths available on the web (Google “bipolar life chart” to find numerous examples), and sophisticated online versions are now available. If the family is willing and able to complete prospective life charts, then the information is highly useful for refining diagnosis and guiding treatment; but the demands of life charting exceed the resources and motivation of many families. The informal “mood and energy checkup” at each visit represents the minimum level of prospective information gathering that should be routine when working with mood disorders.

Measuring Severity

In addition to guiding diagnosis, a second crucial role of assessment tools is quantifying the severity of mood problems. More acute mood disturbance will require a different level of intervention services, with inpatient hospitalization providing the most intensive treatment for severest mood disturbance. Besides navigating the selection of treatment setting, the severity of the mood problems will help prescribe different treatment options. More severe mood problems will suggest the use of pharmacotherapy as a first line treatment, and combination treatments with both psychotherapy and one or more pharmacological agents may be needed to stabilize mood⁹⁹. Moderate levels of mood symptoms may result in lower dose interventions, including outpatient psychotherapy with longer intervals between sessions, or less aggressive dosing of medications. Assessments of severity are also crucial for establishing benchmarks against which to measure treatment response.

Clinician Ratings

Clinical ratings of the severity of mood problems steer the treatment of PBD. Typically the assessment is informal, with all of the attendant limitations described in the discussion of informal diagnosis¹⁰¹. There are global rating scales, such as the Global Assessment of Functioning¹⁷, the CGAS¹¹¹ and the Clinician Global Impressions scale (CGI)¹¹², that assign a number to the clinician’s overall impression of functioning. There also is a bipolar version of the CGI, where the clinician rates manic and depressive symptoms separately⁷⁶.

The next level of sophistication would be to use a semi-structured clinical rating scale. The two most widely used in research with PBD are the Young Mania Rating Scale (YMRS)¹¹³ and the Children’s Depression Rating Scale-Revised (CDRS-R)¹¹⁴. Both have shown evidence of good reliability and acceptable validity in pediatric samples¹¹⁵⁻¹¹⁷. This is reassuring for the YMRS, which was not originally designed for use with children or as an interview¹¹³. The YMRS and CDRS-R both omit symptoms that are DSM-IV criteria for mania or depression, and they omit other associated features that can be important in

assessing the severity of mood disturbance¹⁴. The YMRS does not include a grandiosity item, for example; nor does it measure threat of harm to self. There are newer rating scales designed specifically for use with children and adolescents, that provide more developmentally appropriate anchors, use consistent ratings across all items, and include all DSM symptoms of mood episodes. The KSADS-Mania Rating Scale and Depression Rating Scales have all of these refinements and show good psychometrics¹¹⁸. Clinicians contemplating the use of mood rating scales should be aware that the interviews require a moderate amount of time (typically 15 to 45 minutes), and the semi-structured format is both a blessing and a bane in terms of often leading to sizeable differences in scoring of the same interview by different clinicians⁶⁶.

Checklists: A Reprise

Checklists can complement clinician ratings in measuring severity. Checklists are inexpensive, require virtually no training to use, and incorporate no little or no clinical judgment in scoring; they are the mirror-image of clinician ratings in each of these regards. Checklists also afford the assessment of a broader range of symptom domains than can generally be accomplished via clinician ratings. Parent checklists, in particular, have demonstrated a strong correlation with clinician-rated measures of severity, and they also show good sensitivity to treatment effects^{76,77}.

Quality of Life

There is growing emphasis on quality of life as a vital aspect of the burden of illness and successful outcome of treatment. Several rating scales have been used to measure quality of life in the context of PBD^{119,120}. The KINDL is especially attractive for clinical use because it has two parent and three youth report versions that are developmentally staged, and because it can be used free of charge (<http://www.kindl.org/indexE.html>).

Process Measures to Use During Treatment

Many of the assessment tools discussed above can contribute to process measurement during treatment. Prospective life charts, in-session mood and energy checkups, or repeated administrations of brief rating scales can chart response over the course of intervention. Of the various checklists available, the 10-item versions are generally preferable for repeated administration due to the reduced burden. An exception is the MDQ⁸¹: Although it is a good diagnostic aid, it does not capture information about the severity of current mood problems, so it is not useful as a process or outcome measure.

Prospective life charts also can provide similar information to the three-column and five-column charts used in cognitive behavioral therapy¹²¹. All of these assessment devices ask the patient to chart fluctuations in mood as well as associated events. The three-column chart asks the patient to then also write down what they were thinking at the time, linking the cognition to the emotional response; and the five-column chart goes further by adding an alternative cognition and the emotional response it generates^{121,122}. For those families that are able to do prospective life charting, it is possible to seamlessly weave the components of three- and five-column charting into the recording and in-session discussions of the data.

Other aspects of treatment process that are important to assess include adherence, risk of harm to self and others, and side effects. Measures of adherence can include regularity of kept appointments, completion rates of homework assignments, and compliance with prescribed dosing regimens. Because mood disorder is a major risk factor for suicide and self harm, it is crucial to regularly assess potential suicidal ideation, as well as the presence of a plan and means⁸. Similarly, the degree of aggression and irritability that often manifest with PBD require regular, documented assessment of risk to others. Finally, the potential

side effects for pharmacological treatments are both numerous and potentially quite serious, so they need careful patient education and ongoing monitoring¹²³. Although there are published rating scales for each of these domains, in general they are sufficiently cumbersome that it will usually be more practical to accomplish these goals via direct assessment by the clinician. At the same time, the clinician must follow through on assessing each of these domains and not fall into the trap of irregular assessment and poor documentation.

Measuring Outcomes

Many of the same questionnaires, checklists, and clinician rating scales used to evaluate severity also can provide good measures of outcome. The tools best suited for outcome assessment have good reliability, good content coverage of the relevant symptoms and key aspects of functioning, and are sensitive to treatment effects. Content that is useful for diagnosis may not be identical to the content most useful for outcome measures. For example, irritable mood is ambiguous when used for diagnostic purposes, but irritability is one of the most distressing and impairing features of PBD, and so it definitely merits a central role in outcome assessment. Conversely, two of the eleven items on the YMRS show weak validity when applied to pediatric cases (lack of insight, and bizarre appearance)¹¹⁷, and their inclusion probably dilutes the sensitivity of the YMRS to treatment effects. From a psychometric perspective, outcome measures can and should be longer than the process measures administered more frequently during treatment. The greater length improves reliability and thus the potential validity of the outcome measure¹²⁴, and the length is more likely to be tolerable if administered infrequently. Counter intuitively, statistical power to detect change can be increased by shifting more items to the post-test instead of the pre-test¹²⁵. Clinicians could take advantage of this by using more lengthy outcome measures as part of the termination assessment, where families would not also be spending substantial amounts of time on diagnostic evaluations or administrative paperwork (such as insurance forms). Of the available instruments, the CBCL and the Parent and Adolescent GBI have the most established track record as outcome measures. The CMRS also appears promising. Of the clinician rating scales, the YMRS and CDRS have by far the largest data base in the literature, but the KSADS MRS and DRS deserve consideration due to their advantages in terms of content coverage and developmental appropriateness.

Definitions of Response, Remission, and Clinically Significant Change

In psychiatry outcome studies, treatment response has most commonly been defined using thresholds for percentage reductions in the severity of mood symptoms. For example, a 33% or 50% reduction in YMRS scores from baseline might define “response” to treatment^{126,127}. Such definitions of response are convenient, but they also have some major shortcomings. These limitations include (a) the fact that different patients will need to show varying amounts of change to qualify as a responder, depending upon their initial level of severity; (b) the fact that mania and depression often show different responses to treatment, and sometimes one might worsen at the same time that the other mood symptoms are improving; (c) percentage reductions in symptoms do not necessarily translate into syndromal remission or improvements compared to normative benchmarks; and (d) percentage reductions ignore the degree of precision of an instrument, potentially penalizing more accurate instruments because they are more reliable. The vulnerability of definitions of “treatment response” to unreliability could be a factor contributing to the high rates of placebo response observed in clinical trials.

Multiple refinements have been developed to address these shortcomings of the percentage change approach. These include comparisons to “rules of thumb” about thresholds for mania, hypomania, or depression; the empirical definition of thresholds that distinguish

responders from nonresponders¹²⁸; the use of compound definitions of remission that integrate improvement on mania and depression simultaneously⁷⁶; and the articulation of formal clinical definitions of remission and recurrence¹²⁹. None of these has achieved clear dominance in the research arena, and few have permeated into clinical practice yet.

Perhaps the most fully articulated framework for evaluating outcomes is the “clinically significant change” model proposed by Jacobson and colleagues¹³⁰. Their definition of clinical significant change requires achieving two goals: Demonstrating reliable improvement given the psychometric precision of the outcome measure, and also passing one of three normative benchmarks defined by the range of scores observed in clinical and nonclinical reference samples. Reliable change is tied to the standard error of the difference score, which is a direct measure of the instrument’s precision at measuring change. Jacobson advocated dividing the patient’s raw change score by the standard error of the difference¹³⁰. This would standardize the change score, converting it to a *z*-score (with a mean of zero and standard deviation of one, the same metric as the familiar Cohen’s *d* effect size). Jacobson called this standardized change score the “Reliable Change Index” (RCI). Two advantages of calculating the RCI are (a) that it facilitates comparison of the magnitude of treatment response across different measures (e.g., a 5 point reduction in the YMRS does not mean the same thing as a 5 point reduction in a CDRS-R score; but converting both to RCIs would make clear whether the patient’s mania or depression was responding more to treatment); and (b) that RCIs can be compared to established benchmarks drawn from the normal distribution. RCIs larger than 1.65 are big enough to be 90% certain that the patient is responding, and RCIs larger than 1.96 are enough to be 95% confident. Three drawbacks of the RCI are (a) that it is unfamiliar to most clinicians, (b) it involves computation, and (c) it requires knowledge of the standard error of the difference for the test, which is rarely available. However, all of these problems are tractable. A recent chapter presents the standard error of the difference for several commonly used outcome measures relevant to PBD¹⁴. Table 5 also includes “critical values” expressed in the raw or *T*-score metric that clinicians normally use, as a way of bypassing any computation. For example, if a patient shows at least 10 points of improvement on the YMRS, or 6 points on the CBCL Externalizing, then the clinician can be 90% sure that they are changing in this domain (or 95% sure that the patient is improving, one-tailed).

The second part of the definition of clinically significant change entails demarcating benchmarks based on reference clinical and nonclinical samples. Jacobsen defined three benchmarks, called simply A, B, and C. Youngstrom has suggested referring to them as Away from the clinical range, Back into the normal range, and Closer to the nonclinical than clinical mean¹⁴. The Away threshold is set at two standard deviations below the mean for a clinical sample on the measure¹³⁰. For PBD, estimating the Away threshold would involve finding a sample of youths with PBD and then locating the score on the instrument that fell two standard deviations below the sample mean (roughly corresponding to the 2.5th percentile for cases with PBD). Similarly, the Back into the nonclinical range is established by finding the score that falls two standard deviations above the mean for a nonclinical comparison group. The Closer threshold is found by estimating the weighted mean of the clinical and comparison means, weighting by the sample standard deviations. The technical manuals for many tests include the needed information to calculate the three thresholds, and Cooperberg¹³¹ calculated the thresholds for several other measures pertinent to PBD. Table 5 includes the thresholds for several commonly used outcome measures.

Clinically significant change is frequently used to evaluate individual response in psychotherapy trials. It is a stringent definition of response, and it tends to produce lower estimates of response rates than the simple percentage symptom reduction approach¹³². The conservative definition should cause celebration when actually achieving any of the

definitions of clinically significant change. Studying Table 5 also reveals that some definitions are impossible given the distributional characteristics of a measure. The Away threshold frequently would require obtaining negative scores, because the clinical mean falls within two standard deviations of the lowest possible score on the measure. Similarly, the Back definition would accept high scores (e.g., *T*-scores of 70 on Externalizing or Internalizing problems) as potentially reflecting clinically significant change. This might make sense when coupled with Reliable Change – a seven point reduction in Internalizing Problems from a 76 to a 69 could constitute a substantial improvement, for example. Even so, the thresholds identified by the Closer definition will often provide the most meaningful benchmarks for outcome evaluation¹⁴.

In summary, outcome evaluation is usually done informally, if at all¹⁰¹. However, a variety of assessment tools and definitions of outcome are now available for clinicians working with PBD. Using formal outcome assessment will help better gauge treatment response and enable comparisons between clinical practice and the outcomes described in the literature. The clinically significant change model addresses many of the technical shortcomings of other outcome definitions, and it can now be used with many –but not all—outcome measures relevant to PBD. The necessary benchmarks are not available for the CMRS or KSADS-MRS, for example. On the other hand, all of the different definitions could be applied to commonly used broadband and public domain measures of mania and depression using the information contained in Table 5.

Assessment Tools for Maintenance and Relapse Prevention

Longitudinal data indicate that PBD tends to show a recurrent course^{10,11}, similar to the high rates of relapse observed in adult samples^{8,20}. As treatment progresses, an important component will be planning strategies for monitoring against relapse. Reviewing the life charts and three- or five-column charts would expedite making a list of events likely to trigger exacerbation of mood. There are also normative developmental transitions that are likely to elicit major changes in mood, including the onset of puberty, leaving home, and major role changes such as graduating or failing out of school. Although not all events can be anticipated, many can; and a simple assessment strategy would be to list the likely events and then plan for ways of evaluating mood status when the events arise.

The later phases of treatment can also be a good time to identify warning signs of relapse, or “roughening” of mood¹³³. It is helpful to have manic symptoms under control in order to increase insight into the illness. It also is desirable to have some experience working together with the patient, as many of the cues of relapse are not diagnostic symptoms per se, but rather idiosyncratic aspects of the person’s functioning^{134,135}. A major goal would be to develop warning signs that the patient would use and trust even when getting hypomanic, with the attendant feelings of wellness and loss of insight. An example would be how psychoeducation interventions often help the client learn to use the family or a trusted roommate as a “watchdog” that helps monitor mood¹³⁶.

Given the geographic mobility of both patients and clinicians, a third assessment strategy worth considering is preparing a “care package” for the next clinician as part of the maintenance planning. This care package could be based on a review of the components of treatment, with the patient and practitioner candidly evaluating which were helpful and which were not. Preparing a list of medications tried and responses or adverse events, and a list of therapeutic or lifestyle manipulations and their perceived effectiveness would avoid a lot of guesswork and missteps when resuming active treatment in the aftermath of relapse. Patients and practitioners could keep copies of the documentation to increase the chances that it would be available when needed.

Future Assessment Tools

A great deal of research is investigating methods that could soon contribute to the diagnostic assessment of PBD. Some of the most exciting research includes neurocognitive testing, functional imaging techniques, and genetic testing. In spite of the promise, expert consensus is that none of these methods are currently ready for clinical use. Recent reviews of neurocognitive tests¹³⁷ and neuroimaging^{138,139} have found that the replicated patterns of functioning in PBD tend not to be specific to bipolar disorder, but instead overlap with patterns of functioning seen in ADHD, schizophrenia, and other conditions.

Similarly, the present evidence indicates that bipolar disorder is a polygenic condition, with multiple genes each contributing small increases in the risk of developing the disorder¹⁴⁰. However, several companies now are marketing direct-to-consumer genetic testing, purportedly including tests for bipolar disorder (e.g., <https://psynomics.com/>). This has provoked strong criticism from the academic community, arguing that it is premature to market tests for bipolar disorder and that the results are highly prone to misinterpretation¹⁴¹. The opportunity for misunderstanding is greatest when the results of any of these new methods are treated as yes/no, positive or negative tests for bipolar disorder. This review has demonstrated that although family history, questionnaires, and rating scales are also imperfect measures, once their biases and limitations become known, it is still possible to assimilate them into an evidence-based framework for assessment.

Conclusion

The diagnosis of pediatric bipolar disorder remains controversial in clinical practice. Evidence suggests that it is often diagnosed when not present, and yet many cases of bipolar disorder are also missed. Despite the aura of controversy, there is considerable consensus among experts about the validity of DSM-IV based diagnoses in youth. Marked progress has also been made in validating and honing assessment strategies for PBD. As this review reveals, it is possible to use the literature to inform choices of assessment techniques that contribute to diagnosis, treatment selection, monitoring progress, evaluating outcomes, and long term monitoring and relapse prevention. The review also synthesizes the available literature with the clinical decision-making framework advocated in Evidence Based Medicine and in the clinically significant change literature in psychotherapy research. Both of these frameworks emphasize decision-making about individuals, rather than groups of patients. In consequence, these models speak much more directly to individual clinical care than past research typically has done.

There is no aspect of the assessment of PBD that has been “perfected.” Every assessment tool could be bettered, and each technique offers room for improvement. At the same time, the number of tools and ideas now available for clinicians to apply immediately in their clinical practice is impressively large and diverse. The potential gains from employing evidence based strategies are only hinted at except for in the arena of diagnosis, where recent studies suggest that the contributions could be huge. The goal of clinical research is to answer gaps and uncertainties in clinical practice, and research in PBD is poised to support rapid advances in clinical assessment.

References

1. 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. [PubMed: 17306773]
2. Moreno C, Laje G, Blanco C, et al. National Trends in the Outpatient Diagnosis and Treatment of Bipolar Disorder in Youth. *Archives of General Psychiatry* 2007;64:1032–1039. [PubMed: 17768268]

3. Healy D. The latest mania: selling bipolar disorder. *PLoS Medicine* 2006;3:e185. [PubMed: 16597178]
4. Klein RG, Pine DS, Klein DF. Resolved: Mania is mistaken for ADHD in prepubertal children. *Journal of the American Academy of Child & Adolescent Psychiatry* 1998;37:1093–1096.
5. Faraone SV, Glatt SJ, Tsuang MT. The genetics of pediatric-onset bipolar disorder. *Biological Psychiatry* 2003;53:970–977. [PubMed: 12788242]
6. Geller B, Tillman R. Prepubertal and early adolescent bipolar I disorder: review of diagnostic validation by Robins and Guze criteria. *The Journal of clinical psychiatry* 2005;66(Suppl 7):21–28. [PubMed: 16124838]
7. Youngstrom EA, Birmaher B, Findling RL. Pediatric bipolar disorder: Validity, phenomenology, and recommendations for diagnosis *Bipolar Disorders* 2008;10:194–214.
8. Goodwin, FK.; Jamison, KR. *Manic-depressive illness*. ed 2nd. Oxford University Press; New York: 2007.
9. Perlis R, Miyahara S, Marangell LB, et al. Long-Term implications of early onset in bipolar disorder: data from the first 1000 participants in the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biological Psychiatry* 2004;55:875–881. [PubMed: 15110730]
10. Birmaher B, Axelson D, Strober M, et al. Clinical course of children and adolescents with bipolar spectrum disorders. *Archives of General Psychiatry* 2006;63:175–183. [PubMed: 16461861]
11. Geller B, Tillman R, Bolhofner K, et al. Child bipolar I disorder: prospective continuity with adult bipolar I disorder; characteristics of second and third episodes; predictors of 8-year outcome. *Archives of general psychiatry* 2008;65:1125–1133. [PubMed: 18838629]
12. Lewinsohn PM, Seeley JR, Buckley ME, et al. Bipolar disorder in adolescence and young adulthood. *Child & Adolescent Psychiatric Clinics of North America* 2002;11:461–476. [PubMed: 12222078]
13. Camara, W.; Nathan, J.; Puente, A. Psychological test usage in professional psychology: Report of the APA practice and science directorates. American Psychological Association; Washington, DC: 1998. p. 51
14. Youngstrom, EA. Pediatric bipolar disorder. In: Mash, EJ.; Barkley, RA., editors. *Assessment of Childhood Disorders*. ed 4th. Guilford Press; New York: 2007. p. 253-304.
15. Youngstrom, EA. Evidence-based strategies for the assessment of developmental psychopathology: Measuring prediction, prescription, and process. In: Miklowitz, DJ.; Craighead, WE.; Craighead, L., editors. *Developmental psychopathology*. Wiley; New York: 2008. p. 34-77.
16. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization; London: 1992.
17. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. ed 4th--Text Revision. Author; Washington, DC: 2001.
18. Ghaemi SN, Bauer M, Cassidy F, et al. Diagnostic guidelines for bipolar disorder: a summary of the International Society for Bipolar Disorders Diagnostic Guidelines Task Force Report. *Bipolar disorders* 2008;10:117–128. [PubMed: 18199230]
19. Judd LL, Akiskal HS, Schettler PJ, et al. Psychosocial disability in the course of bipolar I and II disorders: A prospective, comparative, longitudinal study. *Archives of General Psychiatry* 2005;62:1322–1330. [PubMed: 16330720]
20. Kraepelin, E. *Manic-depressive insanity and paranoia*. Livingstone; Edinburgh: 1921.
21. Berk M, Dodd S. Bipolar II disorder: a review. *Bipolar Disorders* 2005;7:11–21. [PubMed: 15654928]
22. Akiskal, HS.; Pinto, O. The soft bipolar spectrum: footnotes to Kraepelin on the interface of hypomania, temperament and depression. In: Marneros, A.; Angst, J., editors. *Bipolar Disorders: 100 years after manic depressive insanity*. Kluwer Academic Publishers, Netherlands; Norwell, MA: 2000. p. 37-62.
23. Youngstrom EA, Youngstrom JK, Starr M. Bipolar Diagnoses in Community Mental Health: Achenbach CBCL Profiles and Patterns of Comorbidity. *Biological Psychiatry* 2005;58:569–575. [PubMed: 15950197]

24. Merikangas KR, Akiskal HS, Angst J, et al. Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry* 2007;64:543–552. [PubMed: 17485606]
25. Grant BF, Stinson FS, Hasin DS, et al. Prevalence, correlates, and comorbidity of bipolar I disorder and axis I and II disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of clinical psychiatry* 2005;66:1205–1215. [PubMed: 16259532]
26. Findling RL, Youngstrom EA, McNamara NK, et al. Early symptoms of mania and the role of parental risk. *Bipolar Disorders* 2005;7:623–634. [PubMed: 16403188]
27. Findling RL, Frazier TW, Youngstrom EA, et al. Double-blind, placebo-controlled trial of divalproex monotherapy in the treatment of symptomatic youth at high risk for developing bipolar disorder. *Journal of Clinical Psychiatry* 2007;68:781–788. [PubMed: 17503990]
28. Kochman FJ, Hantouche E, Ferrari P, et al. Cyclothymic temperament as a prospective predictor of bipolarity and suicidality in children and adolescents with major depressive disorder. *Journal of Affective Disorders* 2005;85:181–189. [PubMed: 15780688]
29. Hantouche EG, Akiskal HS. Toward a definition of a cyclothymic behavioral endophenotype: which traits tap the familial diathesis for bipolar II disorder? *Journal of affective disorders* 2006;96:233–237. [PubMed: 16427137]
30. Klein DN, Depue RA, Slater JF. Inventory identification of cyclothymia. IX. Validation in offspring of bipolar I patients. *Archives of General Psychiatry* 1986;43:441–445. [PubMed: 3964023]
31. Alloy LB, Abramson LY, Walshaw PD, et al. Behavioral Approach System and Behavioral Inhibition System sensitivities and bipolar spectrum disorders: prospective prediction of bipolar mood episodes. *Bipolar disorders* 2008;10:310–322. [PubMed: 18271911]
32. Danielson CK, Youngstrom EA, Findling RL, et al. Discriminative validity of the General Behavior Inventory using youth report. *Journal of Abnormal Child Psychology* 2003;31:29–39. [PubMed: 12597697]
33. Youngstrom EA, Findling RL, Danielson CK, et al. Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. *Psychological Assessment* 2001;13:267–276. [PubMed: 11433802]
34. Depue RA, Slater JF, Wolfstetter-Kausch H, et al. A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: A conceptual framework and five validation studies. *Journal of Abnormal Psychology* 1981;90:381–437. [PubMed: 7298991]
35. Judd LL, Akiskal HS. The prevalence and disability of bipolar spectrum disorders in the US population: re-analysis of the ECA database taking into account subthreshold cases. *Journal of Affective Disorders* 2003;73:123–131. [PubMed: 12507745]
36. Merikangas KR, Herrell R, Swendsen J, et al. Specificity of bipolar spectrum conditions in the comorbidity of mood and substance use disorders: Results from the Zurich cohort study. *Archives of general psychiatry* 2008;65:47–52. [PubMed: 18180428]
37. Axelson DA, Birmaher B, Strober M, et al. Phenomenology of children and adolescents with bipolar spectrum disorders. *Archives of General Psychiatry* 2006;63:1139–1148. [PubMed: 17015816]
38. Joseph M, Youngstrom EA, Soares JC. Antidepressant-Coincident Mania in Children and Adolescents Treated with Selective Serotonin Reuptake Inhibitors. *Future Medicine*. in press.
39. Leibenluft E, Charney DS, Towbin KE, et al. Defining clinical phenotypes of juvenile mania. *The American Journal of Psychiatry* 2003;160:430–437. [PubMed: 12611821]
40. Kowatch RA, Youngstrom EA, Danielyan A, et al. Review and meta-analysis of the phenomenology and clinical characteristics of mania in children and adolescents. *Bipolar Disorders* 2005;7:483–496. [PubMed: 16403174]
41. Youngstrom EA. Definitional issues in bipolar disorder across the life cycle. *Clinical Psychology: Science & Practice*. in press.
42. Stedman JM, Hatch JP, Schoenfeld LS. The current status of psychological assessment training in graduate and professional schools. *Journal of personality assessment* 2001;77:398–407. [PubMed: 11781028]

43. Meyer SE, Carlson GA, Youngstrom E, et al. Long-term outcomes of youth who manifested the CBCL-Pediatric Bipolar Disorder phenotype during childhood and/or adolescence. *Journal of Affective Disorders*. 2008
44. Lewinsohn PM, Klein DN, Seeley J. Bipolar disorder during adolescence and young adulthood in a community sample. *Bipolar Disorders* 2000;2:281–293. [PubMed: 11249806]
45. Tsuchiya KJ, Byrne M, Mortensen PB. Risk factors in relation to an emergence of bipolar disorder: A systematic review. *Bipolar Disorders* 2003;5:231–242. [PubMed: 12895201]
46. Hodgins S, Faucher B, Zarac A, et al. Children of parents with bipolar disorder. A population at high risk for major affective disorders. *Child & Adolescent Psychiatric Clinics of North America* 2002;11:533–553. [PubMed: 12222082]
47. Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. *American journal of medical genetics. Part C, Seminars in medical genetics* 2003;123C:48–58.
48. Youngstrom EA, Duax J. Evidence based assessment of pediatric bipolar disorder, part 1: Base rate and family history. *Journal of the American Academy of Child & Adolescent Psychiatry* 2005;44:712–717. [PubMed: 15968241]
49. Youngstrom EA, Findling RL, Youngstrom JK, et al. Toward an evidence-based assessment of pediatric bipolar disorder. *Journal of Clinical Child and Adolescent Psychology* 2005;34:433–448. [PubMed: 16026213]
50. Strakowski SM, Flaum M, Amador X, et al. Racial differences in the diagnosis of psychosis. *Schizophrenia Research* 1996;21:117–124. [PubMed: 8873779]
51. DelBello MP, Lopez-Larson MP, Soutullo CA, et al. Effects of race on psychiatric diagnosis of hospitalized adolescents: A retrospective chart review. *Journal of Child and Adolescent Psychopharmacology* 2001;11:95–103. [PubMed: 11322750]
52. Duffy A, Alda M, Kutcher S, et al. Psychiatric symptoms and syndromes among adolescent children of parents with lithium-responsive or lithium-nonresponsive bipolar disorder. *The American Journal of Psychiatry* 1998;155:431–433. [PubMed: 9501760]
53. Straus, SE.; Richardson, WS.; Glasziou, P., et al. Evidence-based medicine: How to practice and teach EBM. ed 3rd. Churchill Livingstone; New York: 2005.
54. Meehl, PE. Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. University of Minnesota Press; Minneapolis: 1954.
55. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science* 1989;243:1668–1674. [PubMed: 2648573]
56. Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law* 1996;2:293–323.
57. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 1995;102:684–704.
58. Davidow J, Levinson EM. Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools* 1993;30:351–361.
59. Gigerenzer G. The adaptive toolbox: toward a darwinian rationality. *Nebraska Symposium on Motivation* 2001;47:113–143. [PubMed: 11759346]
60. Lish JD, Dime-Meanan S, Whybrow PC, et al. The National Depressive and Manic-Depressive Association (DMDA) survey of bipolar members. *Journal of Affective Disorders* 1994;31:281–294. [PubMed: 7989643]
61. Hirschfeld RM, Lewis L, Vornik LA. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *Journal of Clinical Psychiatry* 2003;64:161–174. [PubMed: 12633125]
62. Kessler RC, Rubinow DR, Holmes C, et al. The epidemiology of DSM-III-R bipolar I disorder in a general population survey. *Psychological Medicine* 1997;27:1079–1089. [PubMed: 9300513]
63. Zuckerman, M. Vulnerability to psychopathology: A biosocial model. American Psychological Association; Washington, DC: 1999.
64. Youngstrom, JK.; Youngstrom, EA. Bipolar disorder underdiagnosed in community mental health. Biennial Meeting of the International Society of Bipolar Disorders; Delhi, India.

65. Dubicka B, Carlson GA, Vail A, et al. Prepubertal mania: diagnostic differences between US and UK clinicians. *European child & adolescent psychiatry* 2008;17:153–161. [PubMed: 17876503]
66. Mackin P, Targum SD, Kalali A, et al. Culture and assessment of manic symptoms. *British Journal of Psychiatry* 2006;189:379–380. [PubMed: 17012663]
67. Jenkins, M.; Youngstrom, JK.; Perez, Algorta G., et al. How the nomogram improves interpretation of assessment information by clinicians in the community; Annual Convention of the Association for Behavioral and Cognitive Therapy; Orlando, FL. 2008, November;
68. gadoAchenbach, TM.; Rescorla, LA. Manual for the ASEBA School-Age Forms & Profiles. University of Vermont; Burlington, VT: 2001.
69. Gadow, KD.; Sprafkin, J. Child Symptom Inventories Manual. Checkmate Plus; Stony Brook, NY: 1994.
70. Reynolds, CR.; Kamphaus, R. BASC-2 Behavior Assessment System for Children. AGS; Circle Pines, MN: 2004.
71. Mick E, Biederman J, Pandina G, et al. A preliminary meta-analysis of the child behavior checklist in pediatric bipolar disorder. *Biological Psychiatry* 2003;53:1021–1027. [PubMed: 12788247]
72. Kahana SY, Youngstrom EA, Findling RL, et al. Employing parent, teacher, and youth self-report checklists in identifying pediatric bipolar spectrum disorders: an examination of diagnostic accuracy and clinical utility. *Journal of Child and Adolescent Psychopharmacology* 2003;13:471–488. [PubMed: 14977460]
73. Youngstrom EA, Findling RL, Calabrese JR, et al. Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. *Journal of the American Academy of Child & Adolescent Psychiatry* 2004;43:847–858. [PubMed: 15213586]
74. Sackett, DL.; Straus, SE.; Richardson, WS., et al. Evidence-based medicine: How to practice and teach EBM. ed 2nd. Churchill Livingstone; New York: 2000.
75. Scheffer RE, Kowatch RA, Carmody T, et al. Randomized, placebo-controlled trial of mixed amphetamine salts for symptoms of comorbid ADHD in pediatric bipolar disorder after mood stabilization with divalproex sodium. *The American Journal of Psychiatry* 2005;162:58–64. [PubMed: 15625202]
76. Findling RL, McNamara NK, Gracious BL, et al. Combination lithium and divalproex in pediatric bipolarity. *Journal of the American Academy of Child & Adolescent Psychiatry* 2003;42:895–901. [PubMed: 12874490]
77. Findling RL, McNamara NK, Stansbrey R, et al. Combination lithium and divalproex sodium in pediatric bipolar symptom restabilization. *Journal of the American Academy of Child & Adolescent Psychiatry* 2006;45:142–148. [PubMed: 16429084]
78. Miklowitz DJ, Chang KD. Prevention of bipolar disorder in at-risk children: theoretical assumptions and empirical foundations. *Development and psychopathology* 2008;20:881–897. [PubMed: 18606036]
79. Fristad MA. Psychoeducational treatment for school-aged children with bipolar disorder. *Development & Psychopathology* 2006;18:1289–1306. [PubMed: 17064439]
80. Youngstrom EA, Frazier TW, Findling RL, et al. Developing a ten item short form of the Parent General Behavior Inventory to assess for juvenile mania and hypomania. *Journal of Clinical Psychiatry* 2008;69:831–839. [PubMed: 18452343]
81. Wagner KD, Findling RL, Emslie GJ, et al. Validation of the Mood Disorder Questionnaire for Bipolar Disorders in Adolescents. *Journal of Clinical Psychiatry* 2006;67:827–830. [PubMed: 16841633]
82. Youngstrom EA, Meyers OI, Demeter C, et al. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders* 2005;7:507–517. [PubMed: 16403176]
83. Pavuluri MN, Henry DB, Devineni B, et al. Child mania rating scale: Development, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry* 2006;45:550–560. [PubMed: 16601399]
84. Henry DB, Pavuluri MN, Youngstrom E, et al. Accuracy of brief and full forms of the Child Mania Rating Scale. *Journal of Clinical Psychology* 2008;64:368–381. [PubMed: 18302291]

85. Youngstrom EA, Meyers OI, Youngstrom JK, et al. Comparing the effects of sampling designs on the diagnostic accuracy of eight promising screening algorithms for pediatric bipolar disorder. *Biological Psychiatry* 2006;60:1013–1019. [PubMed: 17056395]
86. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Annals of Internal Medicine* 2003;138:40–44. [PubMed: 12513043]
87. Youngstrom EA, Kogos Youngstrom J. Evidence based assessment of pediatric bipolar disorder, part 2: Incorporating information from behavior checklists. *Journal of the American Academy of Child & Adolescent Psychiatry* 2005;44:823–828. [PubMed: 16034285]
88. Sechrest, L.; Davis, MF.; Stickler, TR., et al. Understanding “method” variance. In: Bickman, L., editor. *Research design: Donald Campbell’s legacy*. Sage; Newbury Park, CA: 2000. p. 63-88.
89. Geller B, Warner K, Williams M, et al. 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. [PubMed: 10743842]
90. Hazell PL, Lewin TJ, Carr VJ. Confirmation that Child Behavior Checklist clinical scales discriminate juvenile mania from attention deficit hyperactivity disorder. *Journal of Paediatrics and Child Health* 1999;35:199–203. [PubMed: 10365361]
91. Youngstrom EA, Meyers O, Youngstrom JK, et al. Diagnostic and measurement issues in the assessment of pediatric bipolar disorder: Implications for understanding mood disorder across the life cycle. *Development and Psychopathology* 2006;18:989–1021. [PubMed: 17064426]
92. Loeber R, Green SM, Lahey BB. Mental health professionals’ perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology* 1990;19:136–143.
93. Dell’Osso L, Pini S, Cassano GB, et al. Insight into illness in patients with mania, mixed mania, bipolar depression and major depression with psychotic features. *Bipolar Disorders* 2002;4:315–322. [PubMed: 12479664]
94. Youngstrom EA, Findling RL, Calabrese JR. Effects of adolescent manic symptoms on agreement between youth, parent, and teacher ratings of behavior problems. *Journal of Affective Disorders* 2004;82:S5–S16. [PubMed: 15571790]
95. Youngstrom EA, Joseph MF, Greene J. Comparing the psychometric properties of multiple teacher report instruments as predictors of bipolar disorder in children and adolescents. *J Clin Psychol* 2008;64:382–401. [PubMed: 18300293]
96. Achenbach TM, McConaughy SH, Howell CT. Child/Adolescent behavioral and emotional problems: Implication of cross-informant correlations for situational specificity. *Psychological Bulletin* 1987;101:213–232. [PubMed: 3562706]
97. Carlson GA, Youngstrom EA. Clinical implications of pervasive manic symptoms in children. *Biological Psychiatry* 2003;53:1050–1058. [PubMed: 12788250]
98. Kraemer, HC.; Lowe, KK.; Kupfer, DJ. *To Your Health: How to Understand What Research Tells Us about Risk*. Oxford University Press; New York: 2005.
99. Kowatch RA, Fristad MA, Birmaher B, et al. Treatment guidelines for children and adolescents with bipolar disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 2005;44:213–235. [PubMed: 15725966]
100. Mechanic, D. *Mental Health and Social Policy*. Prentice-Hall, Inc.; Englewood Cliffs, NJ: 1989.
101. Garb, HN. *Studying the clinician: Judgment research and psychological assessment*. American Psychological Association; Washington, DC: 1998.
102. Nottelmann E, Biederman J, Birmaher B, et al. National Institute of Mental Health research roundtable on prepubertal bipolar disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001;40:871–878. [PubMed: 11501685]
103. Kaufman J, Birmaher B, Brent D, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry* 1997;36:980–988. [PubMed: 9204677]
104. Geller B, Zimmerman B, Williams M, et al. 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 & Adolescent Psychiatry* 2001;40:450–455. [PubMed: 11314571]

105. Weller EB, Weller RA, Fristad MA, et al. Children's Interview for Psychiatric Syndromes (ChIPS). *Journal of the American Academy of Child & Adolescent Psychiatry* 2000;39:76–84. [PubMed: 10638070]
106. Geller B, Zimmerman B, Williams M, et al. Phenomenology of prepubertal and early adolescent bipolar disorder: examples of elated mood, grandiose behaviors, decreased need for sleep, racing thoughts and hypersexuality. *Journal of Child and Adolescent Psychopharmacology* 2002;12:3–9. [PubMed: 12014593]
107. Galanter CA, Leibenluft E. Frontiers between attention deficit hyperactivity disorder and bipolar disorder. *Child and adolescent psychiatric clinics of North America* 2008;17:325–346. viii–ix. [PubMed: 18295149]
108. Quinn CA, Fristad MA. Defining and identifying early onset bipolar spectrum disorder. *Current Psychiatry Reports* 2004;6:101–107. [PubMed: 15038912]
109. Masi G, Perugi G, Millepiedi S, et al. Clinical implications of DSM-IV subtyping of bipolar disorders in referred children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* 2007;46:1299–1306. [PubMed: 17885571]
110. Denicoff KD, Smith-Jackson EE, Disney ER, et al. Preliminary evidence of the reliability and validity of the prospective life-chart methodology (LCM-p). *Journal of Psychiatric Research* 1997;31:593–603. [PubMed: 9368200]
111. Shaffer D, Gould MS, Brasic J, et al. A children's global assessment scale (CGAS). *Archives of General Psychiatry* 1983;40:1228–1231. [PubMed: 6639293]
112. National Institute of Mental Health. Rating scales and assessment instruments for use in pediatric psychopharmacology research. *Psychopharmacology Bulletin* 1985;21:839–843.
113. Young RC, Biggs JT, Ziegler VE, et al. A rating scale for mania: Reliability, validity, and sensitivity. *British Journal of Psychiatry* 1978;133:429–435. [PubMed: 728692]
114. Poznanski EO, Miller E, Salguero C, et al. Preliminary studies of the reliability and validity of the Children's Depression Rating Scale. *Journal of the American Academy of Child Psychiatry* 1984;23:191–197. [PubMed: 6715741]
115. 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 & Adolescent Psychiatry* 1992;31:252–257. [PubMed: 1564026]
116. Fristad MA, Weller RA, Weller EB. The mania rating scale (MRS): Further reliability and validity studies with children. *Annals of Clinical Psychiatry* 1995;7:127–132. [PubMed: 8646272]
117. Youngstrom EA, Danielson CK, Findling RL, et al. Factor structure of the Young Mania Rating Scale for use with youths ages 5 to 17 years. *Journal of Clinical Child & Adolescent Psychology* 2002;31:567–572. [PubMed: 12402575]
118. Axelson DA, Birmaher BJ, Brent D, et al. A preliminary study of the Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children mania rating scale for children and adolescents. *Journal of Child and Adolescent Psychopharmacology* 2003;13:463–470. [PubMed: 14977459]
119. Rademacher J, DelBello MP, Adler C, et al. Health-Related Quality of Life in Adolescents with Bipolar I Disorder. *Journal of Child & Adolescent Psychopharmacology* 2007;17:97–103. [PubMed: 17343557]
120. Freeman A, Youngstrom EA, Michalak EE, et al. Quality of life in pediatric bipolar disorder. *Pediatrics*. in press.
121. Danielson CK, Feeny NC, Findling RL, et al. Psychosocial Treatment of Bipolar Disorders in Adolescents: A Proposed Cognitive-Behavioral Intervention. *Cognitive & Behavioral Practice* 2004;11:283–297.
122. Newman, CF.; Leahy, RL.; Beck, AT., et al. *Bipolar disorder: A cognitive therapy approach*. American Psychological Association; Washington, DC: 2002.
123. Wilens, TE. *Straight talk about psychiatric medications for kids*. ed Revised. Guilford Press; New York: 2004.

124. Streiner, DL.; Norman, GR. *Health Measurement Scales: A Practical Guide to their Development and Use.* ed 2nd. Oxford University Press; New York: 1995.
125. Venter, A.; Maxwell, SE. Maximizing Power in Randomized Designs When N is Small. In: Hoyle, RH., editor. *Statistical Strategies for Small Sample Research.* Sage Publications; Thousand Oaks, CA: 1999. p. 31-58.
126. Kafantaris V. Treatment of bipolar disorder in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry* 1995;34:732–741. [PubMed: 7608046]
127. Kowatch RA, Sethuraman G, Hume JH, et al. Combination pharmacotherapy in children and adolescents with bipolar disorder. *Biological Psychiatry* 2003;53:978–984. [PubMed: 12788243]
128. Patel NC, Patrick DM, Youngstrom EA, et al. Response and remission in adolescent mania: Signal detection analyses of the Young Mania Rating Scale. *Journal of the American Academy of Child & Adolescent Psychiatry* 2007;46:628–635. [PubMed: 17450054]
129. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry* 1991;48:851–855. [PubMed: 1929776]
130. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 1991;59:12–19. [PubMed: 2002127]
131. Cooperberg, MD. Characterizing clinical change in youths with bipolar disorders. Case Western Reserve University; United States -- Ohio: 2004.
132. Youngstrom, EA.; Cooperberg, M.; Findling, RL., et al. Identifying the Most Sensitive Outcome Measure for Pediatric Bipolar Disorder. Annual meeting of the American Academy of Child and Adolescent Psychiatry; Miami Beach, FL.
133. Sachs GS. Strategies for improving treatment of bipolar disorder: integration of measurement and management. *Acta Psychiatrica Scandinavica* 2004;(Supplementum):7–17.
134. Jamison, KR. *An Unquiet Mind: A Memoir of Moods and Madness.* Vintage Books; New York: 1995.
135. Hinshaw SP. Parental mental disorder and children's functioning: silence and communication, stigma and resilience. *Journal of Clinical Child and Adolescent Psychology* 2004;33:400–411. [PubMed: 15136205]
136. Miklowitz DJ, Biuckians A, Richards JA. Early-onset bipolar disorder: A family treatment perspective. *Development and Psychopathology* 2006;18:1247–1265. [PubMed: 17064437]
137. Joseph M, Frazier TW, Youngstrom EA, et al. A quantitative and qualitative review of neurocognitive performance in pediatric bipolar disorder. *Journal of Child & Adolescent Psychopharmacology* 2008;18:595–605. [PubMed: 19108664]
138. Chang K, Adleman N, Wagner C, et al. Will neuroimaging ever be used to diagnose pediatric bipolar disorder? *Development and Psychopathology* 2006;18:1133–1146. [PubMed: 17064431]
139. DelBello, MP.; Kowatch, RA. Neuroimaging in pediatric bipolar disorder. In: Geller, B.; DelBello, MP., editors. *Bipolar disorder in childhood and early adolescence.* Guilford; New York: 2003. p. 158-174.
140. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678. [PubMed: 17554300]
141. Braff DL, Freedman R. Clinically responsible genetic testing in neuropsychiatric patients: a bridge too far and too soon. *American Journal of Psychiatry* 2008;165:952–955. [PubMed: 18676598]
142. Pine DS, Cohen P, Gurley D, et al. The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. *Arch Gen Psychiatry* 1998;55:56–64. [PubMed: 9435761]
143. Geller B, Craney JL, Bolhofner K, et al. Two-Year Prospective Follow-Up of Children With a Prepubertal and Early Adolescent Bipolar Disorder Phenotype. *American Journal of Psychiatry* 2002;159:927–933. [PubMed: 12042179]
144. Naylor, MW.; Anderson, TR.; Kruesi, MJ., et al. Pharmacoepidemiology of bipolar disorder in abused and neglected state wards; Poster presented at the National Meeting of the American Academy of Child and Adolescent Psychiatry; San Francisco.

145. Biederman J, Faraone S, Mick E, et al. Attention-deficit hyperactivity disorder and juvenile mania: an overlooked comorbidity? *Journal of the American Academy of Child & Adolescent Psychiatry* 1996;35:997–1008. [PubMed: 8755796]
146. Teplin LA, Abram KM, McClelland GM, et al. Psychiatric Disorders in Youth in Juvenile Detention. *Archives of General Psychiatry* 2002;59:1133–1143. [PubMed: 12470130]
147. Pliszka SR, Sherman JO, Barrow MV, et al. Affective disorder in juvenile offenders: A preliminary study. *The American Journal of Psychiatry* 2000;157:130–132. [PubMed: 10618028]
148. Birmaher B, Ryan ND, Williamson DE, et al. Childhood and adolescent depression: a review of the past 10 years. Part I. *Journal of the American Academy of Child & Adolescent Psychiatry* 1996;35:1427–1439. [PubMed: 8936909]
149. Tillman R, Geller B, Klages T, et al. Psychotic phenomena in 257 young children and adolescents with bipolar I disorder: delusions and hallucinations (benign and pathological). *Bipolar disorders* 2008;10:45–55. [PubMed: 18199241]
150. Achenbach, TM. Manual for the Child Behavior Checklist/4-18 and 1991 profile. University of Vermont, Department of Psychiatry; Burlington: 1991.
151. Achenbach, TM. Manual for the Youth Self Report form and 1991 profile. University of Vermont, Department of Psychiatry; Burlington: 1991.
152. Achenbach, TM. Manual for the Teacher's Report Form and 1991 profile. University of Vermont, Department of Psychiatry; Burlington: 1991.
153. Gracious BL, Youngstrom EA, Findling RL, et al. Discriminative validity of a parent version of the Young Mania Rating Scale. *Journal of the American Academy of Child & Adolescent Psychiatry* 2002;41:1350–1359. [PubMed: 12410078]
154. Hirschfeld RM, Williams JBW, Spitzer RL, et al. Development and validation of a screening instrument for bipolar spectrum disorder: The mood disorder questionnaire. *The American Journal of Psychiatry* 2000;157:1873–1875. [PubMed: 11058490]
155. Tillman R, Geller B. A brief screening tool for a prepubertal and early adolescent bipolar disorder phenotype. *The American Journal of Psychiatry* 2005;162:1214–1216. [PubMed: 15930075]
156. Papolos, DF.; Papolos, J. *The bipolar child: The definitive and reassuring guide to childhood's most misunderstood disorder.* ed 2nd. Broadway Books; New York: 2002.
157. Papolos DF, Hennen J, Cockerham MS, et al. The Child Bipolar Questionnaire: A dimensional approach to screening for pediatric bipolar disorder. *Journal of Affective Disorders* 2006;95:149–158. [PubMed: 16797720]
158. Sackett, DL.; Straus, SE.; Richardson, WS., et al. *Evidence-based medicine: How to practice and teach EBM.* Churchill Livingstone; New York: 1998.
159. Beck, AT.; Steer, RA. *Beck depression inventory manual.* The Psychological Corporation; San Antonio, TX: 1987.

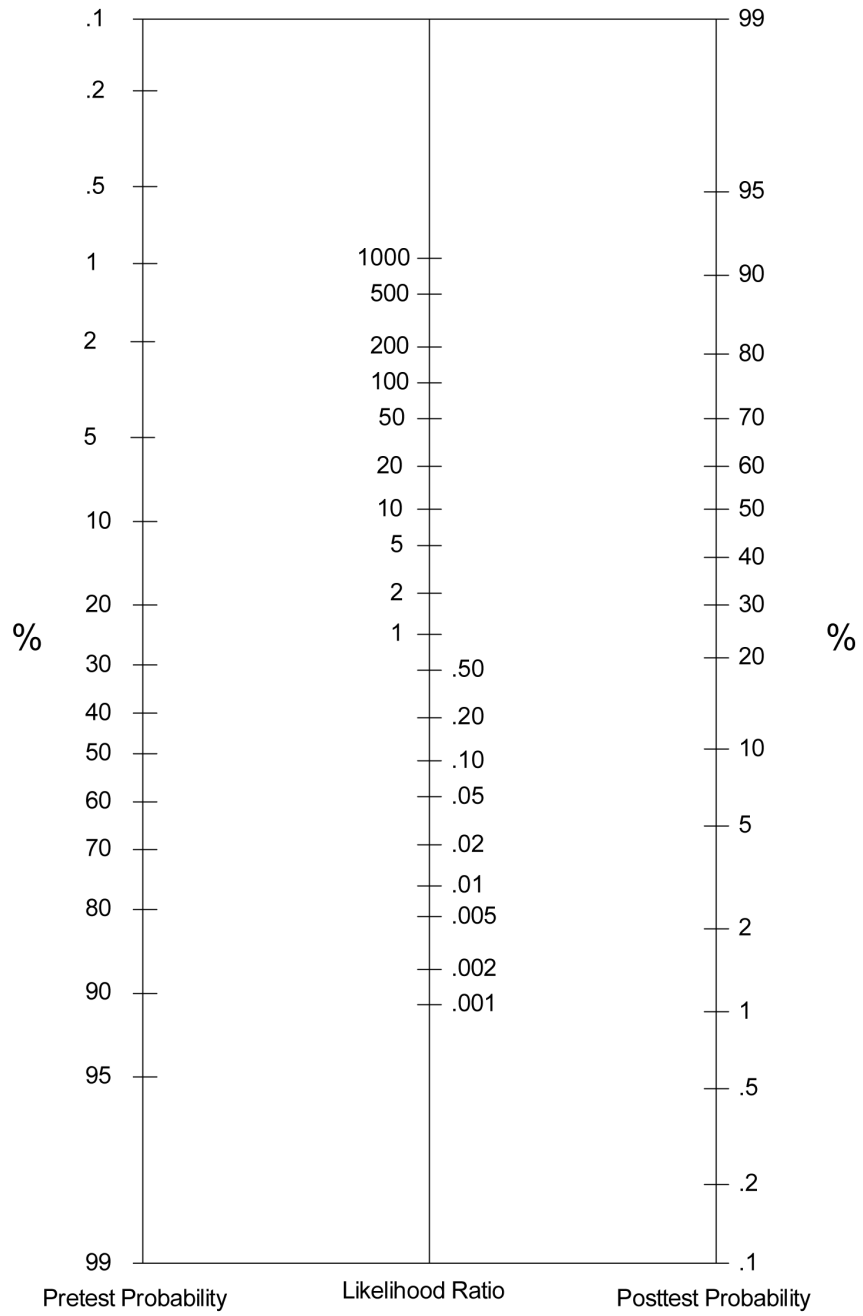


Figure 1.
Nomogram for combining probability with likelihood ratios.

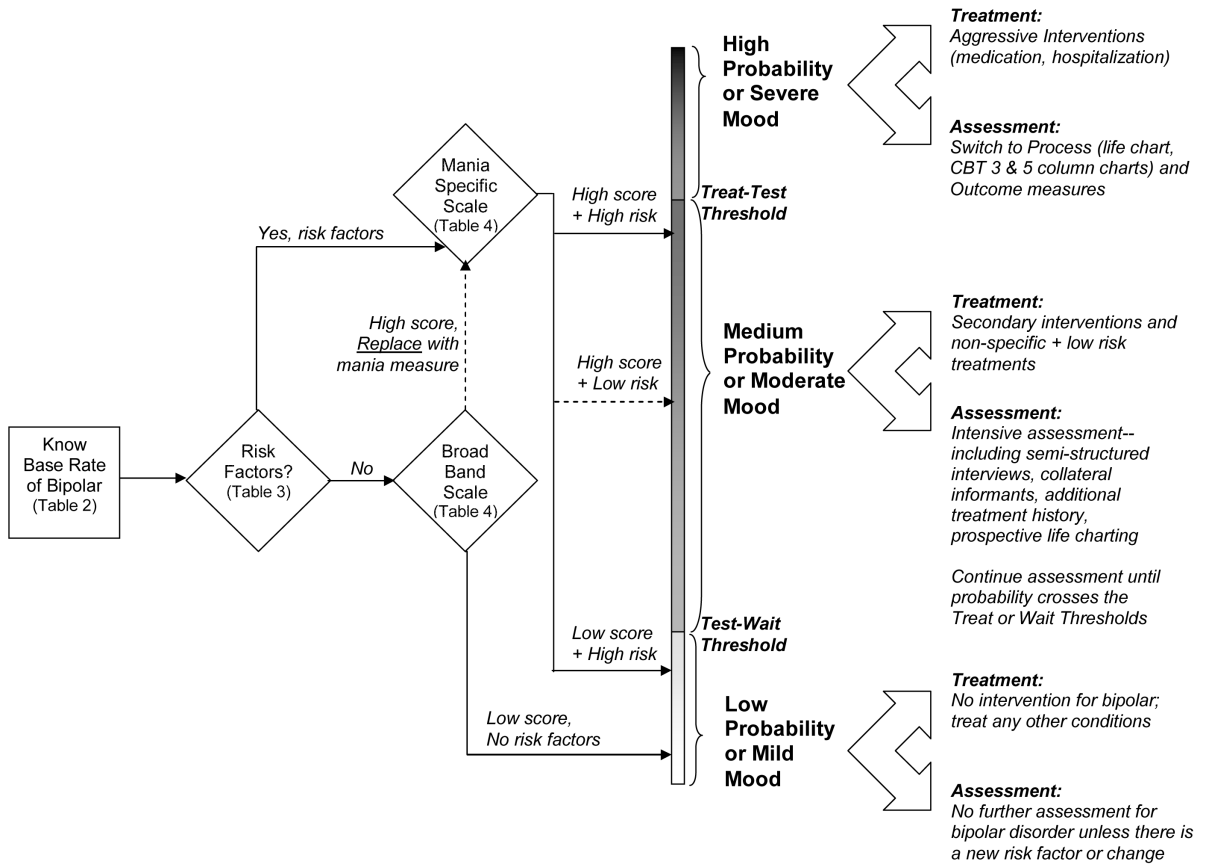


Figure 2.
 Flow chart for assessment and treatment thresholds

Table 1
Diagnoses as a function of past as well as current mood episode, per DSM-IV criteria

Present Episode	Past Episode				
	No Mood	Major Depression	Dysthymic	Hypomanic	Manic/Mixed
No Mood	No mood disorder*	Major depression, single episode*	Dysthymic disorder*	No diagnosis, rule out bipolar NOS*	Bipolar I*
Major Depression	Major depression, in remission	Major depression, recurrent**	Depressive Episode – mild, or in partial remission	Bipolar II	Bipolar I
Dysthymic	Past dysthymic disorder	Major depression, past dysthymic disorder ("Double Depression")	Recurrent dysthymic disorder**	Cyclothymic disorder	Bipolar I
Hypomanic	No mood diagnosis	Bipolar II, current depressed	Cyclothymic disorder, or bipolar NOS	Bipolar NOS (recurrent hypomanias)	Bipolar I
Manic/Mixed	Bipolar I	Bipolar I, current depressed	Bipolar I	Bipolar I	Bipolar I

* Default diagnosis based on accurate identification of the presenting problem without consideration of past mood episodes.

** Note that Kraepelin considered recurrent depressive disorders as potentially on the manic-depressive spectrum, instead of lumping them with the unipolar depressions. We suggest that recurrent depressions, along with early onset depressions, trigger careful evaluation of possible bipolar disorder.

Table 2

Base rates of pediatric bipolar disorder in different settings.

Setting	Base Rate	Demography	Diagnostic Method
General Outpatient Practice ²	1%	All of U.S.A.	National Ambulatory Medical Care Survey
High school epidemiological ⁴⁴	0.6%	Northwestern USA high school	KSADS-PL ^y
Community nonreferred ¹⁴²	1.4%	Upstate New York	
United States epidemiological ²⁴	1.59% child & adolescent onset 4.5% adults	United States General Population	CIDI
United States Child/Adolescent (Lahey et al., unpublished MECA) ⁸	1.8%	Eastern United States	DISC ^{p, y}
Community Mental Health Center ²³	6%	Midwestern Urban, 80% non-white, low-income	Clinical interview & treatment ^{p, y}
General Outpatient Clinic ¹⁴³	6.3%	Urban academic research centers	WASH-U-KSADS ^{p, y}
County Wards (DCFS) ¹⁴⁴	11%	State of Illinois	Clinical interview & treatment ^y
Specialty Outpatient Service ¹⁴⁵	15% to 17%	New England	KSADS-E ^{p, y} (only p young)
Incarcerated Adolescents ¹⁴⁶	2%	Midwestern Urban	DISC ^y
Incarcerated Adolescents ¹⁴⁷	22%	Texas	DISC ^y
Acute psychiatric hospitalizations in 1996 – 2004 – children ¹	26%	All of U.S.A.	Centers for Disease Control survey of discharge diagnoses
Inpatient Service ⁹⁷	30% manic symptoms, <2% strict BP ¹	New York City Metro Region	DICA; KSADS ^{p, y}
Acute psychiatric hospitalizations in 1996 – 2004 – adolescents ¹	34%	All of U.S.A.	Centers for Disease Control survey of discharge diagnoses

^pParent interviewed as component of diagnostic assessment;

^yYouth interviewed as part of diagnostic assessment.

Note. KSADS = Kiddie Schedule for Affective Disorders and Schizophrenia, PL = Present and Lifetime version, WASH-U = Washington University version, -E = Epidemiological version of the KSADS; DSIC = Diagnostic Interview Schedule for Children; DICA = Diagnostic Interview for Children and Adolescents;

CIDI = World Health Organization's Composite International Diagnostic Interview. Table modified from Table 1 in Youngstrom (2007) ¹⁴.

Table 3

Clinical interview red flags that should trigger thorough evaluation of possible PBD.

Red Flag	Description	Reason
Family history of BP*	PBD most likely genetically driven	5x – 10x increase for 1 st degree relative; 2.5x-5x for 2 nd degree relative; 2x for “fuzzy” BP in relative 47,49
Antidepressant Coincident Mania	Manic symptoms while being treated with antidepressants	Undiagnosed PBD 38
Episodic Mood Lability	Rapid switching between depressive and manic symptoms; depressive and manic symptoms at the same time	Common presentation; episodicity more suggestive of mood diagnosis 7
Early Onset Depression	Onset < 25 years	First clinical episode is often depression; substantial portion of pediatric depressions ultimately show bipolar course 28,148
Psychotic features	True delusions/hallucinations in the context of mood	Delusions/Hallucinations common during mood episode 40,149
Episodic Aggressive behavior	Episodic; high-energy. Not instrumental or planned; reactive	Not specific, but common 7,40

* This is the risk factor that is currently most amenable to incorporation in actuarial or Bayesian methods.

Table 4

Areas Under the Curve (AUCs) and Likelihood Ratios for Potential Screening Measures for PBD

Screening Measure	AUC	Diagnostic Likelihood Ratio	Test Score	Clinical Generalizability
Adolescents (11 to 18)				
Parent General Behavior Inventory -- Hypomanic/Biphasic 33	.84 (N=324) 73	.1	<9	Moderate <i>Note: Uses 0 to 3 scoring</i>
		.3	9-15	
		1.1	16-24	
		2.2	25-39	
		4.8	40-48	
		9.2	49+	
CBCL Externalizing T-Score 150	.78 (N=324) 73	.04	<54	Moderate
		.5	54-56	
		1.3	65-69	
		2.1	70-75	
		2.7	76-80	
		4.3	81+	
Parent Mood Disorder Questionnaire 81	~.84 (N<150) 81 .75 (N=124) 82	.3	<5 81	Moderate <i>Note: required co-occurring and at least moderate impairment 81</i>
		3.9	5+	
YSR Externalizing T-Score 151	.71 (N=324) 73	.3	<49	Moderate
		.5	49-55	
		1.2	56-62	
		1.6	63-69	
		2.3	70-76	
		3.0	77+	
TRF Externalizing T-Score 152	.70 (N=324) 73	.3	<46	Moderate
		.6	46-53	
		1.0	54-60	
		2.0	61-68	
		1.5	69-76	
		3.8	77+	
Parent Young Mania Rating Scale 153	.80 (N=324) 73 .70 (N=124) 82	.2	<6 73	High <i>Note: Uses 0 to 4, 0 to 8 scoring 113</i>
		.3	6-11	
		1.0	12-17	
		2.0	18-23	
		4.1	24-27	
		7.4	28+	
Adolescent Mood Disorder Questionnaire 154	~.59 (N< 150) 81 .63 (N=124) 82	.8	<5 81	High <i>Note: Do not include co-occurring and at least moderate impairment items</i>
		1.5	5+	
Adolescent General Behavior Inventory – Hypomanic/Biphasic 34	.62 (N=324) 73 .65 (N=124) 82	.3	<10 73	High <i>Note: Uses 0 to 3 scoring</i>
		1.0	10-17	
		.8	18-26	
		1.2	27-37	
		2.0	38-45	
		3.9	46+	
Adolescent Young Mania Rating Scale -- Questionnaire 82	.50 (N=124) 82	No Better Than Chance		High
Children (5 to 10 years)				
CBCL Externalizing T-Score 150	.82 (N=318) 73	.1	<58	Moderate
		.5	58-67	
		1.5	68-72	
		4.6	73-77	
		3.2	78-81	
		3.5	82+	

Screening Measure	AUC	Diagnostic Likelihood Ratio	Test Score	Clinical Generalizability
Parent General Behavior Inventory -- Hypomanic/Biphasic 33	.81 (N=318) 73	.1	<11	Moderate <i>Note:</i> Uses 0 to 3 scoring
		.5	11-20	
		1.3	21-30	
		2.3	31-42	
		4.9	43-50	
		6.3	51+	
Parent Mood Disorder Questionnaire 81	.72 (N=141) 82	--		High
Parent Young Mania Rating Scale 153	.83 (N=318) 73 .66 (N=141) 82	.1	<7 73	High <i>Note:</i> Uses 0 to 4, 0 to 8 scoring (Young et al., 1978)
		.5	7-13	
		.9	14-21	
		2.8	22-29	
		6.9	30-34	
		8.9	35+	
TRF Externalizing T-Score 152	.57 (N=318) 73	.8	<56	Moderate <i>Note:</i> Not clinically useful
		.9	57-62	
		1.2	63-70	
		1.7	71-77	
		1.3	78+	
Combined Samples (Child & Adolescent Not Reported Separately)				
10 item Parent Child Mania Rating Scale 83,84	.91 (N=150) 84	.2 13.7	<20 20+	Low
Two Item Screen 155	.85 (N=264) 155 .70 (N=500) 85	.31	<9 for 7-8 155	Low 85
			<8 for 9-10	
			<6 for 11+	
		5.2*	9+ for 7-8 8+ for 9-10 6+ for 11+	
General Behavior Inventory Teacher Depression Scale 95	.62	.7 1.3	<4 4+	High
General Behavior Inventory – Teacher Hypomanic/Biphasic Scale 95	Chance	No Better than Chance		High <i>Note:</i> Not clinically useful
Teacher Report Form 152	Chance	No Better than Chance		High <i>Note:</i> Not clinically useful
Child Mania Rating Scale – Teacher 95	Chance	No Better than Chance		High <i>Note:</i> Not clinically useful
Young Mania Rating Scale – Teacher 95	Chance	No Better than Chance		High <i>Note:</i> Not clinically useful
Child Bipolar Questionnaire 156	Not reported (N=135) 157	7.1		Low

Note: All studies used some version of KSADS interview by a trained rater combined with review by a clinician to establish consensus diagnosis. Diagnostic Likelihood Ratio (DLR) refers to the change in probability associated with the test score. Likelihood ratios of 1 indicate that the test result did not change impressions. DLRs larger than 10 or smaller than .10 are frequently clinically decisive; 5 or .2 are helpful, and between 2.0 and .5 are small enough that they rarely result in clinically meaningful changes of formulation¹⁵⁸. Generalizability was rated Low, Moderate, High. Low generalizability meant that the sample was highly selected (e.g., PBD, ADHD, healthy control). Moderate generalizability meant the sample was less highly selected (e.g., few exclusion criteria). High generalizability meant that the sample did not exclude (e.g., took everyone to the clinic).

Table 5
Clinically Significant Change Benchmarks with Common Instruments and Mood Rating Scales

Measure	Cut Scores			
	Away	Back	Closer	90% Critical Change
<i>Benchmarks Based on Published Norms</i>				
<i>Beck Depression Inventory</i> 159				
BDI Mixed Depression	4	22	15	8
<i>CBCL T-Scores (2001 Norms)</i> 68				
Total	49	70	58	4
Externalizing	49	70	58	6
Internalizing	*	70	56	7
Attention Problems	*	66	58	7
<i>TRF T-Scores (2001 Norms)</i> 68				
Total	*	70	57	4
Externalizing	*	70	56	5
Internalizing	*	70	55	7
Attention Problems	*	66	57	7
<i>YSR T-Scores (2001 Norms)</i> 68				
Total	*	70	54	6
Externalizing	*	70	54	8
Internalizing	*	70	54	8
<i>Benchmarks Based on Bipolar Spectrum Samples</i> 131				
Young Mania Rating Scale (Clinician Rated) 113	6	2	2	10
Child Depression Rating Scale-Revised 114	*	40	29	7
Parent GBI – Hypomanic/Biphasic Scale 33	7	19	15	7
Parent GBI – Depression Scale 33	*	18	13	6

Measure	Cut Scores			
	Away	Back	Closer	90% Critical Change
Adolescent GBI – Hypomanic/Biphasic Scale	32 *	32	19	7
Adolescent GBI – Depression Scale	32 *	47	27	9

Note: Away from the clinical range, Back into the nonclinical range, Closer to the nonclinical than clinical mean.

* These benchmarks would require impossible test scores, such as negative numbers.