



Published in final edited form as:

*Psychol Assess.* 2013 March ; 25(1): 12–22. doi:10.1037/a0029225.

## An Inexpensive Family Index of Risk for Mood Issues Improves Identification of Pediatric Bipolar Disorder

**Guillermo Perez Algorta,**

Centro Clinico del Sur, Montevideo, Uruguay, and Department of Psychology, University of North Carolina at Chapel Hill

**Eric A. Youngstrom,**

Departments of Psychology and Psychiatry, University of North Carolina at Chapel Hill & Case Western Reserve University

**James Phelps,**

Department of Psychiatry, Samaritan Mental Health

**Melissa M. Jenkins,**

Department of Psychology, University of North Carolina at Chapel Hill

**Jennifer L. Kogos,** and

Department of Psychology, University of North Carolina at Chapel Hill

**Robert L. Findling**

Department of Psychiatry, Case Western Reserve University School of Medicine

### Abstract

Family history of mental illness provides important information when evaluating pediatric bipolar disorder (PBD). However, such information is often challenging to gather within clinical settings. This study investigates the feasibility and utility of gathering family history information using an inexpensive method practical for outpatient settings. Families ( $N=273$ ) completed family history, rating scales, MINI and KSADS interviews about youths 5–18 (median=11) years presenting to an outpatient clinic. Primary caregivers completed a half page Family Index of Risk for Mood issues (FIRM). All families completed the FIRM quickly and easily. Most (78%) reported 1+ relatives having history of mood or substance issues,  $M=3.7$  ( $SD=3.3$ ). A simple sum of familial mood issues discriminated cases with PBD from all other cases,  $AUROC=.63$ ,  $p=.006$ . FIRM scores were specific to youth mood disorder and not ADHD or disruptive behavior disorder. FIRM scores significantly improved the detection of PBD even controlling for rating scales. No subset of family risk items performed better than the total. Family history information showed clinically meaningful discrimination of PBD. Two different approaches to clinical interpretation showed validity in these clinically realistic data. Inexpensive and clinically practical methods of gathering family history can help to improve the detection of PBD.

### Keywords

pediatric bipolar disorder; family history; community mental health; assessment; screening; sensitivity and specificity

---

Correspondence concerning this article should be sent to: Eric Youngstrom, Department of Psychology, University of North Carolina at Chapel Hill, Davie Hall, CB3270, Chapel Hill, NC 27599–3270. eay@unc.edu.

The other authors have no disclosures.

Bipolar disorder is a highly heritable condition, with both strong genetic (Smoller & Finn, 2003) and environmental contributions (Tsuchiya, Byrne, & Mortensen, 2003) to the risk of illness. Because of this, identifying a family history of mood disorder can be helpful in clarifying the diagnostic formulation for youths (Hodgins, Faucher, Zarac, & Ellenbogen, 2002; Youngstrom & Duax, 2005), who often show ambiguous clinical presentations (Axelson et al., 2006; Lewinsohn, Klein, & Seeley, 2000; Youngstrom, 2009). Family history of bipolar disorder has been recommended as a key piece of evidence to be included in actuarial and *evidence based* approaches for assessing bipolar disorder (Quinn & Fristad, 2004; Youngstrom, Findling, Youngstrom, & Calabrese, 2005). Based on meta-analyses of at-risk youths who have a parent with bipolar disorder, a history of bipolar disorder is associated with at least a five-fold increase in risk for the youth developing bipolar (Hodgins, et al., 2002). Family history of mood disorder—and of bipolar disorder in particular—is useful information for clinicians who are trying to assess risk of bipolar disorder in youths and to weigh and interpret ambiguous clinical presentations. In much the same way, practitioners in other areas of medicine already are using family history, in combination with other established risk factors such as smoking or obesity, to improve clinical assessment and promote early identification of illnesses such as heart disease or cancer (Guyatt & Rennie, 2002).

However, despite the potential utility of family history information, it is often difficult to gather in a systematic fashion. Complicating factors include: a general failure to collect standardized family history as a part of standard practice (Garb, 1998), the expense and cumbersome nature of available semi-structured interviews (Andreasen, Endicott, Spitzer, & Winokur, 1977; Nurnberger et al., 1994; Weissman et al., 2000), the potential for families to be unaware of formal diagnoses or perhaps to have been misdiagnosed (DelBello, Lopez-Larson, Soutullo, & Strakowski, 2001; Neighbors, Trierweiler, Ford, & Muroff, 2003; Strakowski et al., 1997; Strakowski et al., 2003), and the frequent lack of availability of fathers and other relatives for direct interview. These factors often compel clinicians to rely on mothers to provide collateral family history during the evaluation of youths.

There are different strategies to collect information about family history. These approaches can be categorized broadly as the *family history* and *family study* methods. The family history method is a simple report about the presence of specific diseases or various disorders from one family member about another (Andreasen, Endicott, Spitzer, & Winokur, 1977; Andreasen, Rice, Endicott, Reich, & Coryell, 1986; Baker, Berry, & Adler, 1987; Thompson, Orvaschel, Prusoff, & Kidd, 1982). In contrast, the family study method requires the direct clinical assessment of all members of the family. This strategy has higher validity because the diagnosis is more accurate, but it has a markedly higher cost; and it may not be possible for all family members, as some may be unreachable or even deceased (Hardt & Franke, 2007). Whereas the family study method would have the greatest internal validity for research purposes, the family history method more closely approximates what would be typically done in clinical practice, and thus has greater generalizability.

There is a growing consensus in the field that having at least some information about family history is better than not having any (Birmaher et al., 2009; Geller et al., 2006; Hardt & Franke, 2007; Wozniak, Biederman, Mundy, Mennin, & Faraone, 1995). Also, family history about more severe conditions appears to have greater validity than family history about less severe diseases (Hardt & Franke, 2007).

On the other hand, once the data about family history have been collected, the next question is what to do with them? Different scoring strategies have been proposed to optimize the use of this information. Approaches range from a simple dichotomization—family history present/absent—to a more complex scoring mechanism that takes into account the density of

the disorder (i.e., the number of family members who have the disorder; Milne et al., 2008). There is some evidence that density scores have greater predictive validity than the dichotomous score. The observed number of family members with a positive history of disorder is considered the best strategy with disorders with low or moderate prevalence, such as suicide, or bipolar disorder (Milne, et al., 2008).

The goal of the present investigation was to determine the feasibility of gathering family history of mood disorders and related conditions, balancing the competing goals of being clinically meaningful yet sufficiently inexpensive and low-burden to be well tolerated. This study also tested the validity of this index of family history as a risk factor for pediatric bipolar disorder, both by evaluating the discriminative validity of family history as a predictor of youth diagnoses of bipolar disorder, and also by examining the discriminative validity with regard to diagnoses of attention-deficit/hyperactivity disorder (ADHD) in youths. Also, we studied the accuracy of components of the family history index compared to a structured diagnostic interview, a “family study” method of capturing the diagnoses of specific relatives.

Our first hypothesis was that the family history would show predictive value for identifying youths with bipolar spectrum disorder. A second hypothesis was that the association with bipolar disorder would be significantly stronger than for other disorders that are commonly comorbid with bipolar disorder, such as ADHD.

A third hypothesis was that family history, when ascertained using a form that could be readily implemented into clinical practice, would contribute incremental value in the assessment of potential bipolar disorder above and beyond using established mood checklists completed by the same informant.

Finally, we predicted that family history would show significant agreement with diagnostic information collected with a structured diagnostic interview. We predicted only low to moderate kappa values when comparing family history checklist ratings to structured diagnostic interviews about specific relatives, for several reasons: (a) agreement about bipolar diagnoses is typically low when comparing clinical diagnoses to structured diagnostic interviews, with a recent meta-analysis finding  $K < .1$  (Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009); (b) mood diagnoses are especially prone to be misdiagnosed as a psychotic or antisocial disorder in minorities (DelBello, et al., 2001; Neighbors, et al., 2003; Strakowski, et al., 2003), who are over-represented in the present sample, and (c) the risk measure is asking for people’s recall of clinical diagnoses, which is prone to error (Weissman, et al., 2000) and also influenced to an unknown extent by differences in how families conceptualize mood and behavior problems (Li, Silverman, Smith, & Zaccario, 1997).

## Method

### Participants

**Inclusion criteria**—The Institutional Review Board (IRB) of the University Hospitals Case Medical Center and the IRB of Applewood Centers, Inc. (Cleveland, OH) both approved the procedures. Enrolled participants were youth ages 5–18 years old and their primary caregivers seeking outpatient evaluation for the youth. All caregivers gave written informed consent and all youth gave written assent.

**Exclusion criteria**—Families needed to be able to complete questionnaires and interviews in English.

**Demographics and Diagnostic Presentation**—Participants were 273 families presenting for outpatient evaluation of their youth at either an urban community mental health center or an academic outpatient clinic. Families were mostly low-income, with 90% making less than \$40,000 per year, and a median reported income of less than \$15,000 for the primary caregiver. A high school diploma or GED was the median level of education. Seventy five percent of adult informants were biological mothers, 4% were biological fathers, and the remaining 21% of informants consisted of a variety of other relationships, including grandparents (5.9%), aunts or uncles (3.3%), or foster parents (0.4%). Youths were mostly male ( $n = 173$ , 63%), African American ( $n = 187$ , 68%), with an average age of 10.3 ( $SD = 3.6$ ) years.

Diagnostically, 43 youths (16%) were on the bipolar spectrum. Of these, 3 met criteria for bipolar I, 6 for bipolar II, 15 for cyclothymic disorder, and 19 for bipolar NOS. These cases were 56% male, ranged in age from 5 to 17 years ( $M = 10.6$ ,  $SD = 4.0$ ), and racially diverse: 42% identified as African American, 33% as European American, 9% as Hispanic, and 16% as “Other.” Diagnoses that are frequently difficult to discern from pediatric bipolar disorder were highly prevalent in the full sample: 64% of youths had ADHD, 41% had oppositional-defiant disorder (ODD), 31% had a unipolar depressive disorder, and 11% had conduct disorder (CD). In the full sample, the median number of axis I diagnoses was 4.0, and 4.8 in the cases with bipolar disorder.

## Measures

Parents completed a battery of mood and behavior checklists that included the Mood Disorder Questionnaire-Parent version (P-MDQ; Wagner et al., 2006). The P-MDQ is a 13-item scale designed to screening bipolar disorder with a Cronbach’s alpha of .82 in the present sample, and an AUROC = .82 for discriminating youths with bipolar versus all other cases.

Embedded at the end of the P-MDQ was the Family Index of Risk for Mood (FIRM). The FIRM contains a total of 25 checkboxes that consist of an array of questions about mental health history (suicide, depression, mania, hospitalization, or substance use) for each of several relatives (caregiver’s grandparents, parents, aunts/uncles, siblings, or children). The FIRM score consisted of the sum of items endorsed for established risk factors related to bipolar disorder. A copy of the FIRM is appended, and it is available for use by the readership. Separate scores also could count the density of family loading for each type of pathology (i.e., percentage of relatives affected with each type of disorder). Internal consistency, commonly measured by Cronbach’s alpha, does not appear to be a meaningful concept for this type of instrument (Cicchetti et al., 2006). For example, an uncle’s hospitalization would not necessarily be expected to correlate with a sibling’s suicidal ideation.

Parents also completed the 2001 version of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), one of the most widely used instruments in research and clinical work involving child and adolescent mental health. The CBCL includes 118 problem behavior items rated from 0 (*not at all typical of the child*) to 2 (*often typical of the child*). The present study concentrated on the Externalizing Score (8 day test-retest reliability  $r = .92$ , alpha = .94; Achenbach & Rescorla, 2001).

Finally, caregivers also completed the 10 item Mania Scale version of the Parent General Behavior Inventory (PGBI-10M; Youngstrom, Frazier, Demeter, Calabrese, & Findling, 2008). This brief instrument has demonstrated excellent psychometric properties, with a Cronbach’s alpha of .92, a one month retest reliability of .62, and an AUROC = .85 for

discriminating youths with bipolar versus all other cases, similar to an alpha of .93 and AUROC = .83 for the full-length version of the PGBI.

**Youth diagnoses**—Formal diagnoses were made based on an expert review consensus process including the results of an interview using the Schedule for Affective Disorders and Schizophrenia for School-Age Children Present and Lifetime Version (KSADS-PL; Kaufman et al., 1997) supplemented with the mood modules from the Washington University version to gather additional information about mood symptoms and suicidality (Geller et al., 2001). Highly trained research assistants conducted all semi-structured interviews with an item level  $\kappa = 0.85$  (details about training are provided in an earlier preliminary publication; Youngstrom et al., 2005). Interviewers met with the caregiver and the youth sequentially, re-interviewing each as necessary to resolve reporting discrepancies using clinical judgment. A licensed psychologist reviewed the interviews and assigned final consensus diagnoses, blind to scores on the rating scales. Diagnoses followed the DSM-IV criteria (American Psychiatric Association, 2001). Bipolar disorder not otherwise specified (NOS) typically resulted from youths not showing at least one-week durations of mania or four-day durations of hypomanic episodes, rather than having an insufficient number of manic symptoms or low intensity of symptoms. In order to conform with DSM criteria, we did not require elated mood or grandiosity (as would be necessary for the research definition of the *narrow phenotype*; Leibenluft, Charney, Towbin, Bhangoo, & Pine, 2003). However, more than 85% of families reported clear occurrences of one or the other, even though irritable mood and aggression were more commonly perceived as the presenting problem.

**Adult diagnoses**—The relative that brought the youth for evaluation completed a direct interview about their own mental health history, and they repeated the same interview to report on the mental health history of the other biological parent(s) based on The Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). The MINI is a brief, fully-structured diagnostic interview that assesses 17 Axis I disorders, antisocial personality, and suicidality according to DSM-IV criteria. Interviews typically were 15 to 20 minutes per person. The MINI has demonstrated good validity, with median kappas > .63 against other interviews, and inter-rater reliabilities ranging from kappa of .79 to 1.00 (Sheehan et al.).

## Procedure

Families completed the informed consent and assent and then worked with an interview team. One interviewer conducted the KSADS. The other interviewer supervised questionnaire completion and conducted the MINI with the caregiver while the youth was doing the KSADS. Diagnostic interviews were blind to the questionnaire results.

## Data Analyses

Descriptive analyses evaluated distributions against the assumptions for each the proposed analyses. Receiver Operating Characteristic (ROC) analyses quantified the sensitivity and specificity across the full range of possible scores, yielding an Area Under the ROC (AUROC) value where 1.00 would indicate perfect performance and .50 would indicate chance performance of the FIRM when discriminating cases with versus without a bipolar spectrum disorder. A *t* test compared the difference between AUROCS to establish whether one test performed significantly better than the other (Hanley & McNeil, 1983). Logistic regression tested whether the FIRM score provided significant incremental improvement in the prediction of bipolar disorder after controlling for other screening tools. Kappa coefficients quantified the agreement of the FIRM scores about specific relatives with corresponding diagnoses based on the MINI.

## Results

### Descriptive Analyses of the Family Index of Risk for Mood

Families completed the FIRM quickly and without difficulty, despite caregivers' highly variable education levels. Eighty-nine percent completed the FIRM without any questions, and only three needed the instrument read to them. There were virtually no missing data on the FIRM (99.9% complete).

When comparing responses from biological mothers versus all other relatives, children accompanied by their mother tended to be slightly younger ( $p = .04$ ), but showed no other significant demographic or diagnostic differences. Mood scores did not differ significantly either, but biological mothers tended to report more family history of mental health problems than other relatives, ( $t = 2.20, p = .02$ ), consistent with the belief that mothers may be better informed historians than other relatives (Richters, 1992).

The most commonly endorsed family issue in the full sample was "alcohol/drug problems," reported for at least one relative in 62% of families, followed by "depression problems" in 58% of families, "manic or bipolar" in 42% of families, "mental health hospitalization" in 37% of families, and suicide in 23% of the families. Twenty-two percent of families did not endorse even one risk factor. Of the families who endorsed one or more risk factors (78%), the mean number of risk factors endorsed was 3.7 ( $SD = 3.3$ ).

### Hypothesis 1: Family Index of Risk for Mood will be Associated with Pediatric Bipolar Diagnoses

The FIRM total score was significantly higher when the youths had pediatric bipolar diagnoses versus for the rest of families. Except for "alcohol/drug problems," the family risk subscores also were significantly higher in the bipolar group. Effect sizes ranged between Cohen's  $d = .13$  to  $d = .52$ ; see Table 1.

The number of family risk factors (a simple sum of the number of checks) discriminated cases with research diagnoses of pediatric bipolar disorder from all other cases, AUROC = .63,  $p = .006$ . No subset of family risk items performed better than the total. Family history of mania showed essentially identical performance, AUROC = .60,  $p = .035$ .

### Hypothesis 2: Family Index of Risk for Mood will be Specific to Youth Diagnoses of Mood Disorders

The FIRM total score did not show an association with the youth having a diagnosis of ADHD, AUROC = .46,  $p = .355$ ; ODD, AUROC = .50,  $p = .907$ ; or CD, AUROC = .53,  $p = .537$ . The association with bipolar diagnoses was significantly stronger than the association with ADHD, ODD, or CD,  $z$  values  $> 2.3$ ,  $p$  values  $< .01$ . Secondary analyses indicated that the FIRM score was related to unipolar depression in the youths, AUROC = .64,  $p < .0005$ , indicating that the FIRM score reflects risk for mood disorders generally, not just bipolar disorder. When analyses were limited to those with mood disorders, no scales discriminated between youths with unipolar depression versus bipolar disorders.

### Hypothesis 3: FIRM Scores Will Have Incremental Value Above Screening Instruments for Identifying Pediatric Bipolar Disorder

Logistic regressions evaluated whether the FIRM score remained a significant predictor of bipolar diagnoses even after controlling for scores on screening instruments that have previously demonstrated validity in this and other samples. FIRM scores provided a significant improvement in the detection of bipolar cases, whether first adjusting for CBCL Externalizing scores, P-MDQ scores, or PGBI-10M scores (all increments  $p < .05$  for both

FIRM Total and FIRM Mania scores, except for FIRM Total  $p = .070$  after controlling for P-MDQ). The regression weights ranged from .09 for the FIRM after controlling for P-MDQ to .14 after controlling for CBCL Externalizing ( $p = .006$ ), with a one point increase in the FIRM score increasing the predicted odds ratio of the youth having bipolar disorder by 10% to 15% after controlling for the checklist score. Checklist scores were always highly significant, also making a unique incremental contribution to the prediction of bipolar diagnoses. Detailed results are available as supplemental tables upon request from the authors.

### Clinical Decision Making with the Family Index of Risk for Mood

Although logistic regressions provide a good statistical model for evaluating predictors, they are not a practical tool for clinicians to use in evaluating patients (Kraemer, 1992). For this reason, we also evaluated two different approaches for integrating the FIRM into clinical decision making. One approach would be to establish cut scores and report the diagnostic efficiency statistics associated with each. When combining tests -- such as using the FIRM in conjunction with the CBCL, P-MDQ, or PGBI-10M -- the tests can be organized sequentially or in tandem. Because the AUROC values for the FIRM by itself are lower than the AUROC values for both externalizing scores on the CBCL (Youngstrom et al., 2004) and the mania-specific measures such as the P-MDQ and PGBI-10M (Youngstrom et al., 2005), it does not make sense to use the FIRM by itself or as a first line of assessment. Thus we evaluated using FIRM scores as a second, follow-up or in tandem. A strategy of using the family risk variable as a supplemental screening tool, and considering cases “test positive” if they scored high on either the family risk index (scores of 8 or higher) or a mania screen for the youth (e.g., 8 or more on a parent-completed MDQ) resulted in improved diagnostic efficiency, with the algorithm yielding sensitivity of .58 and specificity of .77, ( $LR+ = 2.47$ ,  $LR- = 0.54$ ), and a kappa of .26,  $p < .00005$ . Table 2 presents the diagnostic efficiency statistics for the FIRM total score alone and in combination with either the CBCL Externalizing score using a common “rule of thumb” of  $T > 70$ , or else in tandem with a high score on a mania-specific checklist.

Careful study of Table 2 reveals several things. The number of cases scoring “positive” for bipolar varied widely depending on the algorithm, from 7% to 60% in the present sample. The kappa between the algorithm and the consensus diagnosis was significant for all approaches (except those using the CBCL), but it was also always modest. The percentage of “test positives” that actually have bipolar disorder was never higher than 50% in the present sample, either. The last column in Table 2 uses Bayes’ Theorem to project what the positive predictive value of the algorithm would be if it were used in a different setting where bipolar disorder were more rare. As algebra dictates, making bipolar more rare means that a smaller percentage of test positives will have bipolar disorder, further exacerbating the modest performance of all the algorithms.

The second, newer approach involved estimating diagnostic likelihood ratios for different levels of FIRM scores, which clinicians could then combine with other information about the patient to arrive at a revised estimate of risk of bipolar disorder. This approach has been developed in evidence-based medicine (Guyatt & Rennie, 2002) and has started to be applied to other instruments for assessing pediatric bipolar disorder (Youngstrom, et al., 2004; Youngstrom & Youngstrom, 2005). In the simple case where the base rate of bipolar disorder is the only prior information available about risk and the FIRM score is the only piece of information added, then the combination of these data points -- whether via Bayes’ Theorem or a probability nomogram -- is the positive predictive value. This approach is more flexible than the older, multiple-test sequencing approach. Using the “multi-level” approach, where likelihoods are estimated for multiple ranges of scores, milks more information from the test result than a simple “high/low” dichotomization. The approach also allows

combinations of tests that may not yet have been empirically evaluated together, and it also enables projections of scenarios that will occur in clinical practice but may be too rare to empirically examine with parametric statistics.

Table 3 presents the multi-level likelihood ratios for splitting FIRM scores into “low,” “moderate,” and “high risk” scores, and then illustrates the resulting values when combining these with a high score on a specific test (e.g., high score on the PGBI-10M). The pairing of a high risk FIRM score *and* a high risk PGBI-10M score yielded an estimate of 69% probability that the youth has a bipolar diagnosis, versus the closest analog estimate from Table 2 being a 33% probability for a high FIRM score *or* a high PGBI-10M. The use of an “or” strategy will always be less specific than an “and” strategy, allowing more false positives (Guion, 1998; Youngstrom, Findling, & Calabrese, 2003). However, trying to evaluate the “and” strategy using the multi-test sequencing approach would run aground as the sample became too shallow to explore the combination of interest: only 7 cases scored high on both measures, failing to meet Kraemer’s rule of thumb for evaluating a medical test (Kraemer, 1992).

### Validity Analyses

We studied the criterion validity of the scores collected from parents with the FIRM (bipolar, unipolar depression, alcohol and substance abuse) as compared to MINI *family study method* findings about relatives’ diagnoses. The kappa between parent’s FIRM and MINI mania or hypomania was  $K = .23$  ( $p < .00005$ ); for depression,  $K = .26$  ( $p < .00005$ ). For alcohol and substance abuse, kappas were .24 and .21 ( $p < .00005$ ) respectively. When the two approaches disagreed, the MINI identified more cases of bipolar than did the FIRM by a ratio of 2.1 to 1, indicating that the FIRM was more specific than sensitive.

### Discussion

The goal of this paper was to evaluate the clinical feasibility and utility of a short checklist to gather information about familial risk for bipolar disorder. Based on the literature about the lag in recognition of bipolar disorder (Hirschfeld, Lewis, & Vornik, 2003; Lish, Dime-Meenan, Whybrow, Price, & Hirschfeld, 1994; Marchand, Wirth, & Simon, 2006) and the frequency with which it goes undiagnosed or misdiagnosed, particularly in minorities (DelBello, et al., 2001; Strakowski, et al., 2003), the tool included items assessing related characteristics beyond the DSM-IV criteria for depression and mania. The brief family history items were well tolerated by families, who answered all items and had little to no difficulty with the reading level and organization of questions. When item scores pertaining to bipolar, depression, and substance use were compared to the results of structured diagnostic interviews for the same relatives, the FIRM showed modest sized but highly significant kappas, consistent with the typical performance of brief family history measures compared to direct interviews (Hardt & Franke, 2007; Roy, Walsh, & Kendler, 1996; Weissman, et al., 2000). Also consistent with other measures, the FIRM was more likely to omit cases identified by direct structured interview than to have false positives.

More importantly, the family history information showed a clinically meaningful association with youth diagnoses of pediatric bipolar disorder (based on strict DSM-IV criteria, and applied via a semi-structured diagnostic interview conducted by highly trained raters). The association between family history and diagnosis appeared specific to mood disorders, and was not associated with changes in risk of ADHD or disruptive behavior disorders. The value of the FIRM score appeared similar for identifying those at risk of mood disorders generally, rather than bipolar disorder specifically, although developing a clinical interpretative framework for predicting depression falls outside the scope of this paper. Results were consistent with the general pattern of findings from twin studies, where mood



disorders show distinct heritability contributions from externalizing problems (Rende & Waldman, 2006) or substance disorders (Kendler et al., 1995). The size of the relationship is also comparable to established benchmarks based on reviews of studies looking at familial risk (DelBello & Geller, 2001; Hodgins, et al., 2002): The diagnostic likelihood ratio of 2.5 for high scores on the FIRM is similar to the risk associated with confirmed bipolar disorder in a second-degree relative, or a fuzzy history of bipolar (Youngstrom, et al., 2005).

In addition, the family history information provided incremental validity when predicting bipolar diagnoses, even after controlling for other information provided by the same informant. These analyses provided a strong test of the potential clinical value of adding the FIRM to other assessment strategies. It also is worth noting that these results were found in a sample that contained many characteristics likely to challenge a test's performance. The entire sample had serious enough problems to be seeking services, with high degrees of comorbidity in both the youths and their families. The diagnoses most difficult to tease apart from bipolar (Kim & Miklowitz, 2002) outnumbered the number of cases with bipolar disorder. Furthermore, the majority of the bipolar cases had "spectrum" presentations that often slip past screening tools (Miller, Klugman, Berv, Rosenquist, & Ghaemi, 2004) yet appear to be the more common presentation according to epidemiological studies (Merikangas & Pato, 2009; Van Meter, Moreira, & Youngstrom, 2011). Whereas effect sizes typically shrink when moving from "efficacy" research designs that emphasize internal validity into "effectiveness" settings that emphasize generalizability, the present findings are "pre-shrunk" to the extent that the design incorporated many of the factors that would be typically encountered in clinical applications.

It was interesting to find that the risk index did not improve as a predictor of pediatric bipolar disorder when limited to family history of mania. This could be due to bipolar disorder resulting from the accumulation of multiple nonspecific risk factors (Tsuchiya, et al., 2003), or else due to the inaccuracy with which bipolar disorder has been recognized in the past. This could be error in past diagnoses, or it could be the product of the mental health literacy (Jorm, 2000) of the caregiver responsible for completing the FIRM. Overall, the findings suggest that even inexpensive and highly simplified methods of gathering family history can help to improve the detection of pediatric bipolar disorder.

Finally, we also investigated how the FIRM might be applied by clinicians, either alone or in combination with other rating scales. We evaluated both a multiple-test sequence and a newer, likelihood-ratio/Bayesian approach advocated by evidence based medicine. Comparison of the two showed that the newer method is more flexible, gleaning more information from the same tests than a simple "test positive/negative" decision, allowing more choice in terms of test selection, and allowing projections to cases encountered in clinical practice. These projections will not be perfect, and should be updated or superseded as new data become available; but the Bayesian framework also provides a structure for integrating these updates (Smith, Winkler, & Fryback, 2000) and for generating reasonable estimates with imperfect inputs (Straus, Richardson, Glasziou, & Haynes, 2005). Using these approaches is likely to improve the accuracy of decisions about diagnoses (Rettew, et al., 2009), particularly about bipolar disorder in youths (Jenkins, Youngstrom, Washburn, & Youngstrom, 2011). Our recommendation to clinicians would be to combine the FIRM with whatever general intake assessment that they use, and combine the risk information from it and any other risk factors or assessment scales using the nomogram approach to decide whether the patient is low, medium, or high risk of bipolar disorder (Youngstrom, Freeman, & Jenkins, 2009). Further assessment and treatment formulation would then proceed accordingly.

## Limitations and Strengths

As mentioned above, one of the main limitations is that the present sample includes many demographic and clinical characteristics that are likely to reduce the diagnostic performance of the FIRM. It is likely that the performance of the FIRM would be different, and potentially even better, in samples with a different composition (Zhou, Obuchowski, & McClish, 2002). Test developers often use designs that create optimal performance for the measure (Tillman & Geller, 2005), but the performance of these instruments can degrade rapidly under clinically realistic conditions (Youngstrom, Meyers, Youngstrom, Calabrese, & Findling, 2006). It also is possible that a more complicated scoring algorithm, using customized weights for different relatives or varying clinical issues, might further improve performance of the FIRM (Milne et al., 2008). However, these weights are also more likely to be sample-dependent and to shrink upon cross-validation or application in clinical settings. Most importantly, any family history measure is limited by the knowledge of the informant. For example adopted children, or mothers who are unaware of the paternal side of the family, will not have the same historical information available. Also, lack of a *reported* family history does not equate to lack of a family history, due to all of the factors that can undermine the validity of any one person's knowledge of a given family's history.

## Clinical Implications and Future Directions

Future research should study how the FIRM and the interpretive approach might apply to other clinical issues, such as depression or ADHD. Studies should also investigate the extent to which education or cultural factors might change the performance of the FIRM, as well as the role of other factors such as family conflict as predictors in their own right. It is reassuring that other evidence-based assessment recommendations have remained robust when generalized to new demographic groups and clinical settings (Jenkins, Youngstrom, Youngstrom, Feeny, & Findling, 2011). Another important angle of study would be whether different family members agree when completing the FIRM, and whether it is possible to select which perspective would have the greatest informational value (Vandeleur et al., 2008).

Present results suggest that the FIRM could be applied as part of a comprehensive assessment approach for pediatric bipolar disorder. It is low cost and low burden enough to be practical in most clinical settings, and it has demonstrated incremental value even under clinically realistic conditions. A vignette included in the appendices illustrates how the FIRM score might be integrated with other information within this evidence-based medicine framework to support flexible but accurate evaluation of bipolar disorder in youths. Although a direct family interview would be more accurate (and would yield more powerful information), the FIRM is user friendly and stands a good chance of being implemented in settings where a direct interview may not be possible. On the other hand, the FIRM is not a good proxy for direct interviews of family members when family history is the main focus, consistent with the findings for other family history screens (Li et al., 1997). Clinicians who are familiar with genograms may want to draw one with the family before asking the parent to complete the FIRM, as this process has increased the yield of useful family history information in other studies (Baker, et al., 1987).

## Acknowledgments

We thank the families who participated in this research. This work was supported in part by NIH R01MH066647 (PI: E. Youngstrom). E. Youngstrom has acted as a consultant with Lundbeck and received travel support from Bristol-Myers Squibb. Dr. Findling receives or has received research support, acted as a consultant, received royalties from, and/or served on a speaker's bureau for Abbott, Addrenex, Alexza, American Psychiatric Press, AstraZeneca, Biovail, Bristol-Myers Squibb, Forest, GlaxoSmithKline, Guilford Press, Johns Hopkins University Press, Johnson & Johnson, KemPharm Lilly, Lundbeck, Merck, National Institutes of Health, Neuropharm,

Novartis, Noven, Organon, Otsuka, Pfizer, Physicians' Post-Graduate Press, Rhodes Pharmaceuticals, Roche, Sage, Sanofi-Aventis, Schering-Plough, Seaside Therapeutics, Sepracore, Shionogi, Shire, Solvay, Stanley Medical Research Institute, Sunovion, Supernus Pharmaceuticals, Transcept Pharmaceuticals, Validus, WebMD and Wyeth. Dr. Phelps discontinued participation on speakers' bureaus in 2008; prior to that time he received honoraria from GlaxoSmithKline and AstraZeneca. He receives royalties from McGraw-Hill for a book on Bipolar II in adults.

## References

- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont; 2001.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 2001. Text Revision ed
- Andreasen N, Endicott J, Spitzer R, Winokur G. The family history method using diagnostic criteria: reliability and validity. *Archives of General Psychiatry*. 1977; 34:1229.10.1001/archpsyc.1977.01770220111013 [PubMed: 911222]
- Andreasen N, Rice J, Endicott J, Reich T, Coryell W. The family history approach to diagnosis: How useful is it? *Archives of General Psychiatry*. 1986; 43:421.10.1001/archpsyc.1986.01800050019002 [PubMed: 3964020]
- Axelson DA, Birmaher B, Strober M, Gill MK, Valeri S, Chiappetta L, Keller M. Phenomenology of children and adolescents with bipolar spectrum disorders. *Archives of General Psychiatry*. 2006; 63:1139–1148.10.1001/archpsyc.63.10.1139 [PubMed: 17015816]
- Baker NJ, Berry SL, Adler LE. Family diagnoses missed on a clinical inpatient service. *American Journal of Psychiatry*. 1987; 144:630–632. [PubMed: 3578573]
- Birmaher B, Axelson D, Monk K, Kalas C, Goldstein B, Hickey MB, Brent D. Lifetime psychiatric disorders in school-aged offspring of parents with bipolar disorder: the Pittsburgh Bipolar Offspring study. *Archives of General Psychiatry*. 2009; 66:287–296.10.1001/archgenpsychiatry.2008.546 [PubMed: 19255378]
- Cicchetti D, Bronen R, Spencer S, Haut S, Berg A, Oliver P, Tyrer P. Rating scales, scales of measurement, issues of reliability: resolving some critical issues for clinicians and researchers. *Journal of Nervous and Mental Disease*. 2006; 194:557–564.10.1097/01.nmd.0000230392.83607.c5 [PubMed: 16909062]
- DelBello MP, Geller B. Review of studies of child and adolescent offspring of bipolar parents. *Bipolar Disorders*. 2001; 3:325–334.10.1034/j.1399-5618.2001.30607.x [PubMed: 11843782]
- DelBello MP, Lopez-Larson MP, Soutullo CA, Strakowski SM. Effects of race on psychiatric diagnosis of hospitalized adolescents: A retrospective chart review. *Journal of Child and Adolescent Psychopharmacology*. 2001; 11:95–103.10.1089/104454601750143528 [PubMed: 11322750]
- Garb, HN. Studying the clinician: Judgment research and psychological assessment. Washington, DC: American Psychological Association; 1998.
- Geller B, Tillman R, Bolhofner K, Zimmerman B, Strauss NA, Kaufmann P. Controlled, blindly rated, direct-interview family study of a prepubertal and early-adolescent bipolar I disorder phenotype: morbid risk, age at onset, and comorbidity. *Archives of General Psychiatry*. 2006; 63:1130–1138.10.1001/archpsyc.63.10.1130 [PubMed: 17015815]
- Geller B, Zimmerman B, Williams M, Bolhofner K, Craney JL, DelBello MP, Soutullo C. Reliability of the Washington University in St. Louis Kiddie Schedule for Affective Disorders and Schizophrenia (WASH-U-KSADS) mania and rapid cycling sections. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2001; 40:450–455.10.1097/00004583-200104000-00014 [PubMed: 11314571]
- Guion, RM. Assessment, measurement, and prediction for personnel decisions. Hillsdale, NJ: Erlbaum; 1998.
- Guyatt, GH.; Rennie, D., editors. Users' guides to the medical literature. Chicago: AMA Press; 2002.
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983; 148:839–843. [PubMed: 6878708]

- Hardt J, Franke P. Validity, reliability and objectivity of the family history method in psychiatry: A meta analysis. *European Psychiatry*. 2007; 22:49–58.10.1016/j.eurpsy.2006.10.003 [PubMed: 17188848]
- 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; 64:161–174.10.4088/JCP.v64n0209 [PubMed: 12633125]
- Hodgins S, Faucher B, Zarac A, Ellenbogen M. Children of parents with bipolar disorder. A population at high risk for major affective disorders. *Child & Adolescent Psychiatric Clinics of North America*. 2002; 11:533–553.10.1016/S1056-4993(02)00002-0 [PubMed: 12222082]
- Jenkins MM, Youngstrom EA, Washburn JJ, Youngstrom JK. Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology: Research and Practice*. 2011; 42:121–129.10.1037/a0022506
- Jenkins MM, Youngstrom EA, Youngstrom JK, Feeny NC, Findling RL. Generalizability of evidence-based assessment recommendations for pediatric bipolar disorder. *Psychological Assessment*. 2011 epub ahead of print. 10.1037/a0025775
- Jorm AF. Mental Health Literacy: public knowledge about mental disorders. *British Journal of Psychiatry*. 2000; 177:396–401.10.1192/bjp.177.5.396 [PubMed: 11059991]
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Ryan N. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1997; 36:980–988.10.1097/00004583-199707000-00021 [PubMed: 9204677]
- Kendler KS, Walters EE, Neale MC, Kessler RC, Heath AC, Eaves LJ. The structure of the genetic and environmental risk factors for six major psychiatric disorders in women. Phobia, generalized anxiety disorder, panic disorder, bulimia, major depression, and alcoholism. *Archives of General Psychiatry*. 1995; 52:374–383.10.1001/archpsyc.1995.03950170048007 [PubMed: 7726718]
- Kim EY, Miklowitz DJ. Childhood mania, attention deficit hyperactivity disorder and conduct disorder: a critical review of diagnostic dilemmas. *Bipolar Disorders*. 2002; 4:215–225.10.1034/j.1399-5618.2002.01191.x [PubMed: 12190710]
- Kraemer, HC. *Evaluating medical tests: Objective and quantitative guidelines*. Newbury Park, CA: Sage Publications; 1992.
- Leibenluft E, Charney DS, Towbin KE, Bhangoo RK, Pine DS. Defining clinical phenotypes of juvenile mania. *American Journal of Psychiatry*. 2003; 160:430–437.10.1176/appi.ajp.160.3.430 [PubMed: 12611821]
- Lewinsohn PM, Klein DN, Seeley J. Bipolar disorder during adolescence and young adulthood in a community sample. *Bipolar Disorders*. 2000; 2:281–293.10.1034/j.1399-5618.2000.20309.x [PubMed: 11249806]
- Li G, Silverman JM, Smith CJ, Zaccario ML. Validity of the family history method for identifying schizophrenia-related disorders. *Psychiatry Research*. 1997; 70:39–48.10.1016/S0165-1781(97)03120-x [PubMed: 9172275]
- Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RM. The National Depressive and Manic-Depressive Association (DMDA) survey of bipolar members. *Journal of Affective Disorders*. 1994; 31:281–294.10.1016/0165-0327(94)90104-X [PubMed: 7989643]
- Marchand WR, Wirth L, Simon C. Delayed diagnosis of pediatric bipolar disorder in a community mental health setting. *Journal of Psychiatric Practice*. 2006; 12:128–133.10.1097/00131746-200603000-00011 [PubMed: 16728912]
- Merikangas KR, Pato M. The epidemiology of bipolar disorder in adults and children: Magnitude, correlates and future directions. *Clinical Psychology: Science & Practice*. 2009; 16:121–133.10.1111/j.1468-2850.2009.01152.x
- Miller CJ, Klugman J, Berv DA, Rosenquist KJ, Ghaemi SN. Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder. *Journal of Affective Disorders*. 2004; 81:167–171.10.1016/S0165-0327(03)00156-3 [PubMed: 15306144]

- Milne BJ, Moffitt TE, Crump R, Poulton R, Rutter M, Sears MR, Caspi A. How should we construct psychiatric family history scores? A comparison of alternative approaches from the Dunedin Family Health History Study. *Psychological Medicine*. 2008; 38:1793.10.1017/s0033291708003115 [PubMed: 18366822]
- Neighbors HW, Trierweiler SJ, Ford BC, Muroff JR. Racial Differences in DSM Diagnosis Using a Semi-Structured Instrument: The Importance of Clinical Judgment in the Diagnosis of African Americans. *Journal of Health and Social Behavior*. 2003; 44:237–256.10.2307/1519777 [PubMed: 14582306]
- Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Reich T. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Archives of General Psychiatry*. 1994; 51:849–859.10.1001/archpsyc.1994.03950110009002 [PubMed: 7944874]
- Quinn CA, Fristad MA. Defining and identifying early onset bipolar spectrum disorder. *Current Psychiatry Reports*. 2004; 6:101–107.10.1007/s11920-004-0049-1 [PubMed: 15038912]
- Rende, R.; Waldman, I. Behavioral & molecular genetics and developmental psychopathology. In: Cicchetti, D.; Cohen, DJ., editors. *Developmental Psychopathology: Developmental Neuroscience*. 2. Vol. 2. Hoboken, NJ: Wiley; 2006. p. 427-464.
- Retzew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research*. 2009; 18:169–184.10.1002/mpr.289 [PubMed: 19701924]
- Richters JE. Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin*. 1992; 112:485–499.10.1037//0033-2909.112.3.485 [PubMed: 1438639]
- Roy M-A, Walsh D, Kendler KS. Accuracies and inaccuracies of the family history method: A multivariate approach. *Acta Psychiatrica Scandinavica*. 1996; 93:224–234.10.1111/j.1600-0447.1996.tb10639.x [PubMed: 8712019]
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Dunbar GC. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*. 1998; 59:22–33.10.4088/JCP.09m05305whi [PubMed: 9881538]
- Smith JE, Winkler RL, Fryback DG. The first positive: computing positive predictive value at the extremes. *Annals of Internal Medicine*. 2000; 132:804–809. 200005160-00008. [PubMed: 10819704]
- 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:48–58.10.1002/ajmg.c.20013
- Strakowski SM, Hawkins JM, Keck PE Jr, McElroy SL, West SA, Bourne ML, Tugrul KC. The effects of race and information variance on disagreement between psychiatric emergency service and research diagnoses in first-episode psychosis. *Journal of Clinical Psychiatry*. 1997; 58:457–463.10.4088/JCP.v58n1010a [PubMed: 9375599]
- Strakowski SM, Keck PE Jr, Arnold LM, Collins J, Wilson RM, Fleck DE, Adebimpe VR. Ethnicity and diagnosis in patients with affective disorders. *Journal of Clinical Psychiatry*. 2003; 64:747–754.10.4088/JCP.v64n0702 [PubMed: 12934973]
- Straus, SE.; Richardson, WS.; Glasziou, P.; Haynes, RB. *Evidence-based medicine: How to practice and teach EBM*. 3. New York: Churchill Livingstone; 2005.
- Thompson W, Orvaschel H, Prusoff B, Kidd K. An evaluation of the family history method for ascertaining psychiatric disorders. *Archives of General Psychiatry*. 1982; 39:53.10.1001/archpsyc.1982.04290010031006 [PubMed: 7055408]
- Tillman R, Geller B. A brief screening tool for a prepubertal and early adolescent bipolar disorder phenotype. *American Journal of Psychiatry*. 2005; 162:1214–1216.10.1176/appi.ajp.162.6.1214 [PubMed: 15930075]
- Tsuchiya KJ, Byrne M, Mortensen PB. Risk factors in relation to an emergence of bipolar disorder: A systematic review. *Bipolar Disorders*. 2003; 5:231–242.10.1034/j.1399-5618.2003.00038.x [PubMed: 12895201]

- Van Meter A, Moreira AL, Youngstrom EA. Meta-analysis of epidemiological studies of pediatric bipolar disorder. *Journal of Clinical Psychiatry*. 2011; 72:1250–1256.10.4088/JCP.10m06290 [PubMed: 21672501]
- Vandeleur CL, Rothen S, Jeanpratre N, Lustenberger Y, Gamma F, Ayer E, Preisig M. Inter-informant agreement and prevalence estimates for substance use disorders: Direct interview versus family history method. *Drug and Alcohol Dependence*. 2008; 92:9–19.10.1016/j.drugalcdep.2007.05.023 [PubMed: 17643870]
- Wagner KD, Hirschfeld R, Findling RL, Emslie GJ, Gracious B, Reed M. Validation of the Mood Disorder Questionnaire for Bipolar Disorders in Adolescents. *Journal of Clinical Psychiatry*. 2006; 67:827–830.10.4088/JCP.v67n0518 [PubMed: 16841633]
- Weissman MM, Wickramaratne P, Adams P, Wolk S, Verdelli H, Olfson M. Brief screening for family psychiatric history: the family history screen. *Archives of General Psychiatry*. 2000; 57:675–682.10.1001/archpsyc.57.7.675 [PubMed: 10891038]
- Wozniak J, Biederman J, Mundy E, Mennin D, Faraone SV. A pilot family study of childhood-onset mania. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1995; 34:1577–1583.10.1097/00004583-199512000-00007 [PubMed: 8543528]
- Youngstrom E, Meyers O, Demeter C, Youngstrom J, Morello L, Piiparinen R, Findling RL. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders*. 2005; 7:507–517.10.1111/j.1399-5618.2005.00269.x [PubMed: 16403176]
- Youngstrom E, Van Meter A, Algorta GP. The bipolar spectrum: myth or reality? *Current Psychiatry Reports*. 2010; 12:479–489.10.1007/s11920-010-0153-3 [PubMed: 20857347]
- Youngstrom EA. Definitional issues in bipolar disorder across the life cycle. *Clinical Psychology: Science & Practice*. 2009; 16:140–160.10.1111/j.1468-2850.2009.01154.x
- Youngstrom EA, Duax J. Evidence-based assessment of pediatric bipolar disorder, part I: base rate and family history. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2005; 44:712–717.10.1097/01.chi.0000162581.87710.bd [PubMed: 15968241]
- Youngstrom EA, Findling RL, Calabrese JR. Who are the comorbid adolescents? Agreement between psychiatric diagnosis, parent, teacher, and youth report. *Journal of Abnormal Child Psychology*. 2003; 31:231–245. [PubMed: 12774858]
- Youngstrom EA, Findling RL, Calabrese JR, Gracious BL, Demeter C, DelPorto Bedoya D, Price M. Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2004; 43:847–858.10.1097/01.chi.0000125091.35109.1e [PubMed: 15213586]
- Youngstrom EA, Findling RL, Youngstrom JK, Calabrese JR. Toward an evidence-based assessment of pediatric bipolar disorder. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:433–448.10.1207/s15374424jccp3403\_4 [PubMed: 16026213]
- Youngstrom EA, Frazier TW, Demeter C, Calabrese JR, Findling RL. Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. *Journal of Clinical Psychiatry*. 2008; 69:831–839. ej07m03819. [PubMed: 18452343]
- Youngstrom EA, Freeman AJ, Jenkins MM. The assessment of children and adolescents with bipolar disorder. *Child & Adolescent Psychiatry Clinics of North America*. 2009; 18:353–390.10.1016/j.chc.2008.12.002
- 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; 44:823–828.10.1097/01.chi.0000164589.10200.a4 [PubMed: 16034285]
- Youngstrom EA, Meyers OI, Demeter C, Kogos Youngstrom J, Morello L, Piiparinen R, Calabrese JR. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders*. 2005; 7:507–517.10.1111/j.1399-5618.2005.00269.x [PubMed: 16403176]
- Youngstrom EA, Meyers OI, Youngstrom JK, Calabrese JR, Findling RL. Comparing the effects of sampling designs on the diagnostic accuracy of eight promising screening algorithms for pediatric

bipolar disorder. *Biological Psychiatry*. 2006; 60:1013–1019.10.1016/j.biopsych.2006.06.023 [PubMed: 17056395]

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

### Appendix I. Family Index of Risk for Mood (FIRM)

	Please indicate whether any of your (blood) relatives have had any of these concerns:					other than the child in this study
	Grandparents	Parents	Aunts/Uncles	Brothers/Sisters	Children	
Suicide	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Alcohol/Drug Problems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Mental Hospital	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Depression Problems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Manic or Bipolar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Has a health professional ever told you that you have manic-depressive illness or bipolar disorder?						Yes No

### Appendix II Vignette Illustrating the Use of the Family Index of Risk for Mood as Part of a Clinical Evaluation

Lena is a 12 year old African American girl who was evaluated in a community mental health center for concerns about her social and emotional functioning. She has been doing more poorly in school this past fall and is extremely irritable and argumentative at home. In order to gain some context for the problems that worried her family and the school staff, Lena’s mother sought an outpatient mental health evaluation. As part of the standard intake procedure, the mother completed the CBCL Achenbach scale and the Family Index of Risk for Mood issues (FIRM), the brief family screen described in the article. The CBCL indicated a Tscore of 70 on the Externalizing Problem Scale, reflecting a clinically elevated level of aggressive and rule-breaking behavior compared to other girls of similar age. The total FIRM score was 12, due to heavy family history of severe problems -- including suicide, bipolar and drug/alcohol history in an uncle.

At this early stage, the clinician has not spent any additional time with the family, nor have they added any additional assessment tools or evaluations. There are three key pieces of information relevant to Lena’s probability of having a bipolar spectrum disorder: (a) that her problems are bringing her to an outpatient clinic; (b) the elevated CBCL Externalizing score; and (c) the FIRM score. Depending on the setting, between 5 and 15% of new referrals to outpatient mental health clinics are likely to be on the bipolar spectrum. The clinician elects to start with a 6% probability, based on published recommendations and on their recent pattern of referrals. The clinician decides to use the recommended evidence-based medicine procedure – a probability nomogram – to integrate the initial screening results. Based on published benchmarks, the CBCL Externalizing score increases the likelihood of a bipolar disorder by 1.5 times, and the FIRM score increases the likelihood by 2.5 times. Combining these pieces of information using a probability nomogram (<http://www.cebm.net/index.aspx?o=1043>) (Guyatt & Rennie, 2002; Youngstrom & Duax, 2005; Youngstrom, Freeman, & Jenkins, 2009) yields a combined probability of 19% -- bipolar disorder may not be likely, but there are warning signs that warrant further investigation. Diagnostic likelihood ratios are changes in the odds of a diagnosis, not linear changes in probability of the diagnosis. The probability nomogram saves the clinician several steps when compared to

calculating the change in probability directly. The algebraic steps involved would be (a) convert the prior probability to prior odds; (b) multiply the odds by the diagnostic likelihood ratio of the test or risk factor, and then (c) converting the revised odds back into a probability. When more than one diagnostic likelihood ratio is available simultaneously, it is more convenient to multiply the diagnostic likelihood ratios and then enter the product in the nomogram or calculator, versus iterating through the steps sequentially with each likelihood ratio; algebraically the final result will be the same.

The clinician decides to have the mother complete a specialized mania scale, the Parent General Behavior Inventory-Mania 10 Item version (Youngstrom, et al., 2008). This is also brief and in the public domain, again taking little time and adding no cost to the evaluation. The score comes back a 19, highly elevated. Consulting with the benchmarks shows that this increases the likelihood of a bipolar disorder by 7.5, substantially more worrisome than the score on the CBCL. Recommended practice is to focus on the single most relevant score from any rating scales gathered from the same person. Thus the PGBI-10M replaces the CBCL for the purpose of evaluating potential bipolar disorder. The clinician then combines the base rate of bipolar in outpatient settings (6%) with the likelihoods attached to high FIRM (2.5) and high PGBI-10M scores (7.5). Using a nomogram or probability calculator arrives at an estimate of 54% revised probability of a bipolar spectrum disorder. This alerts the clinician that detailed evaluation of the possibility of a bipolar disorder is justified, although the available information is not sufficient to justify pharmacological intervention without further assessment. At this stage, inexpensive screening tools have helped identify risk factors and focus attention on priorities for further assessment.

The clinician reviews the FIRM results in detail with the mother and learns that Lena's father and grandmother suffered unipolar depression and substance abuse problems in the past, and one of Lena's brothers is actually in treatment after being diagnosed as having bipolar II. The clinician chooses to replace the information about the family history from the FIRM score with the information about the bipolar II in the brother. A confirmed history of bipolar disorder in a first-degree relative is linked with at least a 5.0 increase in likelihood. Consulting the nomogram one last time results in a revised probability of 70% (6% base rate combined with 7.5 likelihood from the PGBI-10M, and 5.0 likelihood from the brother's bipolar II diagnosis).

This example illustrates how information can be integrated, and rapid choices made about how to upgrade information and re-evaluate, without adding much time or expense to existing procedures. At this point, a direct discussion can be had about the costs and benefits of different treatment options and more intensive assessment strategies. In Lena's case, a careful semi-structured interview revealed that she met criteria for DSM-IV diagnoses of cyclothymic disorder and comorbid ADHD. Lena and her family agreed to begin psychotherapy as a first line strategy, focused on mood monitoring, emotion regulation, and family-focused therapy (Youngstrom, Van Meter, & Algorta, 2010). A difficult decision remains to be solved about the incorporation of a pharmacological strategy for Lena's ADHD. The family agreed to a stimulant trial in combination with a daily life chart to track Lena's mood and energy while also monitoring potential side effects.



**Table 1**

Family Index of Risk for Mood issues (FIRM) scores for families of youths with bipolar spectrum diagnoses compared to rest of outpatient mental health sample

FIRM Scale	Non- Bipolar ( <i>n</i> =230) <i>M</i> ( <i>SD</i> )	Bipolar Spectrum ( <i>n</i> =43) <i>M</i> ( <i>SD</i> )	Mann Whitney <i>z</i> value	Cohen's <i>d</i>
<i>Suicide</i>	0.2 (0.5)	0.5 (0.7)	2.71 **	.44
<i>Alcohol/Drug</i>	1.1 (1.2)	1.3 (1.2)	.95	.13
<i>Mental Hospitalizations</i>	0.4 (0.7)	0.7 (1.0)	2.07 *	.41
<i>Depression</i>	1.0 (1.1)	1.6 (1.4)	2.72 **	.49
<i>Manic</i>	0.6 (0.9)	1.1 (1.3)	2.37 **	.49
<i>Total Score</i>	3.4 (3.2)	5.2 (3.8)	2.77 **	.52

Note.

\*  
 $p < .05$

\*\*  
 $p < .01$

\*\*\*  
 $p < .005$ ; significance tests based on nonparametric Mann-Whitney because of high skew; Cohen's *d* based on pooled *SD*s; conventional benchmarks for *d* are .2 for small, .5 for medium, and .8 for large effects.

**Table 2**

Diagnostic Efficiency of Different Algorithms Screening for Pediatric Bipolar Disorder

Algorithm	Sensitivity	Specificity	Test Positive Rate	Kappa	PPV	Projected PPV (5% prevalence)
FIRM Score 8+	.28	.89	.14	.18**	.32	.12
P-MDQ 8+	.49	.84	.21	.28***	.49	.14
CBCL Ext 70+	.63	.42	.59	.02	.15	.05
PGBI-10M 18+	.19	.95	.07	.18**	.40	.17
FIRM 8+ or P-MDQ 8+	.58	.77	.29	.26***	.31	.07
FIRM 8+ or CBCL 70+	.65	.65	.60	.03	.17	.09
FIRM 8+ or PGBI-10M 18+	.40	.40	.18	.23***	.33	.13

Note.

\*  $p < .05$

\*\*

$p < .01$

\*\*\*

$p < .005$ ; "test positive rate" refers to the percentage of the sample that would "test positive" based on each algorithm. It is also known as the "level" of the test in Kraemer's (1992) terminology.

**Table 3**  
Change in Odds of Bipolar Diagnosis (Diagnostic Likelihood Ratios) Based on Family Index of Risk for Mood (FIRM) Scores

FIRM Total Score	Descriptor	Sample %	Likelihood Ratio for Bipolar	Likelihood Ratio for Mood	Posterior Probability of Bipolar (5% prevalence)	Probability Combined with High Score on PGBI-10M (18+)
0 to 2	Low Risk	45%	0.6	0.6	.03	.33
3 to 7	Moderate Risk	42%	1.1	1.1	.05	.47
8 or more	High Risk	14%	2.5	4.3	.12	.69

*Note.* The sample used to evaluate the diagnostic efficiency was an outpatient mental health sample with 16% prevalence of bipolar spectrum disorder and 100% of youths meeting criteria for some kind of diagnosis. PPV = Positive Predictive Value.