



NIH PUBLIC ACCESS

Author Manuscript

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2015 April 03.

Published in final edited form as:

J Clin Child Adolesc Psychol. 2014 ; 43(4): 552–565. doi:10.1080/15374416.2014.883930.

Clinical Decision-Making about Child and Adolescent Anxiety Disorders Using the Achenbach System of Empirically Based Assessment

Anna Van Meter,

Yeshiva University, Albert Einstein College of Medicine, Ferkauf Graduate School of Psychology

Eric A. Youngstrom,

Departments of Psychology and Psychiatry, University of North Carolina at Chapel Hill

Thomas Ollendick,

Child Study Center, Department of Psychology, Virginia Polytechnic Institute and State University

Christine Demeter, and

Department of Psychiatry, Case Western Reserve University School of Medicine

Robert L. Findling

Department of Psychiatry, Johns Hopkins University, and Kennedy Krieger Institute

Abstract

Objective—Anxiety disorders are common among children, but can be difficult to diagnose. An actuarial approach to the diagnosis of anxiety may improve the efficiency and accuracy of the process. The objectives of this study were to determine the clinical utility of the Achenbach CBCL and YSR, two widely used assessment tools, for diagnosing anxiety disorders in youth, and to aid clinicians in incorporating scale scores into an actuarial approach to diagnosis through a clinical vignette.

Method—Demographically diverse youth, aged 5 to 18 years, were drawn from two samples; one ($N=1084$) was recruited from a research center, the second ($N=651$) was recruited from an urban community mental health center. Consensus diagnoses integrated information from semi-structured interview, family history, treatment history, and clinical judgment.

Results—The CBCL and YSR internalizing problems T scores discriminated cases with any anxiety disorder or with GAD from all other diagnoses in both samples (p values $<.0005$); the two scales had equivalent discriminative validity (p values $>.05$ for tests of difference). No other scales, nor any combination of scales, significantly improved on the performance of the Internalizing scale. In the highest risk group, Internalizing scores >69 (CBCL) or >63 (YSR) resulted in a Diagnostic Likelihood Ratio of 1.5; low scores reduced the likelihood of anxiety disorders by a factor of 4.

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. 216-410-7975 cell, 919-962-2537 fax.

The other authors have no disclosures.

Conclusions—Combined with other risk factor information in an actuarial approach to assessment and diagnosis, the CBCL and YSR Internalizing scales provide valuable information about whether or not a youth is likely suffering from an anxiety disorder.

Keywords

anxiety; evidence based; children and adolescents; assessment; diagnosis

Assessment and diagnosis guide case conceptualization and treatment. Childhood disorders are difficult to diagnose: Confounding factors – developmental stage, family constellation, school environment, comorbid psychiatric disorders or physical illnesses –render few cases “by the book.” Anxiety disorders may be particularly difficult to diagnose, in part because some degree of anxiety is developmentally appropriate for children (Sakolsky & Birmaher, 2008). Although it can be tempting to adopt a “wait and see” philosophy with these cases, no one wants to make children and parents suffer needlessly if effective treatment is available. Further, untreated anxiety disorders in childhood are likely to lead to chronic mental health problems (Pauschardt, Remschmidt, & Mattejat, 2010). However, if the symptoms are due to an issue other than anxiety, whether it be depression, a medical condition, or a difficult social situation, one would not want to administer inappropriate treatment. High rates of comorbidity among youth with anxiety complicate the diagnostic picture further (Aschenbrand, Angelosante, & Kendall, 2005).

Anxiety disorders are relatively common among children, with lifetime prevalence rates estimated between 9–20% (Aschenbrand et al., 2005; Kessler et al., 2005; Merikangas, He, Brody, et al., 2010; Merikangas, He, Burstein, et al., 2010; Sakolsky & Birmaher, 2008). However, community prevalence is not necessarily a good indicator of the frequency with which clinicians will see youth with anxiety disorders. The prevalence rate of anxiety disorders will shift depending on the clinical environment and geographic location, among other factors. Knowing how often one should *expect* to see anxiety disorders is an important first step in formulating accurate diagnoses based on data (Meehl & Rosen, 1955; Straus, Glasziou, Richardson, & Haynes, 2011; Youngstrom, 2013).

Taking a data-driven approach to diagnosis aligns with the push to incorporate evidence-based practice into child psychology and psychiatry (Chambless & Ollendick, 2001), and, specifically, into diagnostic assessment methods (Cohen et al., 2008). Evidence-based assessment is consistently more accurate than clinical decision making as usual (Grove, 1987; Jenkins, Youngstrom, Washburn, & Youngstrom, 2011; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). The choices made regarding the design of an assessment protocol should promote progress toward at least 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 family (Youngstrom, 2008). The Three P framework reduces the use of extraneous assessment tools, which unnecessarily increase burden and cost and can blur the diagnostic picture by introducing irrelevant information (Kraemer, 1992).

How does one incorporate assessment data into a diagnosis? Most often, practitioners rely on their clinical judgment, weighing their diagnostic impressions, along with test scores and

other factors, to come to a decision (Garb, 1998). This is a complicated process with a “black box” feel to it. Clinical diagnoses have remarkably low reliability when compared to each other or to structured diagnostic interviews (Rettew et al., 2009). Evidence-Based Medicine (EBM) (Straus et al., 2011) recommends using validated assessment tools, along with an actuarial approach to diagnostic decision-making (Dawes, Faust, & Meehl, 1989; Meehl, 1954; Straus & McAlister, 2000).

The EBM method relies on combining the available facts, such as prevalence rate, family history, and scores on validated measures, to determine the probability that a child has a particular disorder. It helps clinicians to make sense of what they know about their patients, and it does so in a consistent and reliable way. There are a number of methods one can use to combine the probabilities within a Bayesian framework, including online tools and mobile phone apps (Straus, Tetroe, & Graham, 2011). An alternative that does not require computation or software is the probability nomogram (see Figure 1), which is an easy, paper-and-pencil tool for revising diagnostic probabilities (Straus et al., 2011). The nomogram is flexible, providing an estimate of the likelihood that an individual meets criteria for a specific disorder (known as posterior probability) by synthesizing available information, which the clinician can then use in case formulation. Unlike the DSM diagnostic scales produced by many questionnaires, an EBM approach does not equate a positive test with a diagnosis. Instead, the EBM framework integrates the change in risk attached to a test score with other key information, to yield a single, integrated probability estimate (Youngstrom, 2013). Included at the end of this paper is a vignette, in which we illustrate how the nomogram can be used in clinical practice.

Clinical interviews are time consuming, and there is an inherent tension between reliability and burden, with structured and semi-structured approaches often increasing the duration of the interview, but unstructured approaches often producing poor reliability (Garb, 1998; Rettew et al., 2009). Questionnaires are easier to validate in regard to their diagnostic ability, and can be completed more quickly than a full diagnostic interview (Aschenbrand et al., 2005a). The Achenbach System of Empirically Based Assessment is one of the most widely used assessment tools in child psychology and psychiatry (Achenbach, 2000; Pauschardt et al., 2010). It is popular among both clinicians and researchers, making it more likely than other questionnaires to inform an EBA approach (Achenbach, 2005).

Previous studies have found that the CBCL and its counterpart, the Youth Self Report (YSR; Achenbach, 1991b), can frequently identify anxiety disorders (Aschenbrand et al., 2005; Ferdinand, 2008; Pauschardt, Remschmidt, & Mattejat, 2010; Warnick, Bracken, & Kasl, 2008).

However, results of previous studies have been mixed (Warnick et al., 2008), and findings have not been presented in a way that makes it easy for clinicians to incorporate the data in an evidence-based assessment approach. Furthermore, the CBCL and YSR comprise a number of potentially relevant subscales including the Total Problems score, Internalizing and Externalizing scores, Anxious/Depressed, Withdrawn/Depressed, Somatic, Social Problems, Thought Problems, Attention, and DSM scales for Affective Disorders and Anxiety Disorders. A previous study (Pauschardt et al., 2010) found that the DSM-oriented

Anxiety Disorders CBCL subscale was the best at predicting any anxiety disorder, with an Area Under the Curve (AUC) of .71. It was the only scale with at least “medium” discriminative ability, per Swets’ (1988) benchmarks (low=0.5–0.7; medium=0.7–0.9; high >0.9). Most scores produced from the CBCL offer, at best, low discriminative ability. This is surprising considering that several CBCL scales measure anxiety symptoms. Interestingly, in another study, Pauschardt et al (2010) found that the DSM-oriented Anxiety Problems CBCL subscale had very poor internal consistency, drawing into question its reliability. In contrast, Ebesutani et al. (2010) found that the CBCL DSM-oriented Anxiety Problems scale was good at discriminating separation anxiety disorder, generalized anxiety disorder, and specific phobia from both patients without anxiety disorders and youth with mood disorders (all AUCs>0.80). The Anxious/Depressed scale also had moderate discriminative validity against mood disorders (AUC=0.72) and non-anxiety disorders (AUC=0.80).

Previous studies have focused on fairly homogenous populations; most often white youth presenting to outpatient, specialty anxiety clinics. Given that the discriminative ability of the CBCL, even among these samples, has been inconsistent, it is crucial to know how the CBCL and YSR perform in demographically and diagnostically heterogeneous samples that would be more generalizable to a broad range of clinical settings. The present study uses large samples from two populations. The first group, recruited from an outpatient academic clinic, was similar to the samples from previous studies of the CBCL and anxiety disorders. The second, from an urban community mental health clinic, was composed of youth from primarily low-income, minority families; most had comorbid disorders, particularly externalizing disorders, and their families were often naïve to mental health services (Youngstrom et al., 2005). Including this second group enables us to test whether the findings from the academic, research clinic would generalize to an applied, clinical setting, chosen *a priori* to have markedly different demographics and referral patterns. To prevent the interviewer from being a confound, all of the interviewers involved in the community mental health setting also saw families at the academic clinic. This design allowed us to compare the discriminative validity of the CBCL across samples and to determine whether demographics or clinical features moderated the scales’ diagnostic validity. Consistent performance would reinforce the generalizability of the results, whereas significant differences would generate hypotheses about potential moderators.

Based on findings from earlier studies (Aschenbrand et al., 2005a; Ferdinand, 2008; Pauschardt et al., 2010), we expected the CBCL and YSR to show statistical validity, significantly discriminating cases with anxiety from other diagnoses, and we expected the diagnostic efficiency (e.g., AUC) to be better for *any* anxiety disorder than for specific anxiety disorders. Additionally, we hypothesized that both caregiver and youth report would be significantly more discriminating than teacher report on the same scales (Youngstrom et al., 2005). We expected the CBCL and YSR both to perform better in the outpatient research clinic sample than in the community mental health clinic, due to the demographic differences and clinical complexity of the community mental health setting. Finally, we estimated multilevel likelihood ratios (Jaeschke, Guyatt, & Sackett, 1994) for ranges of scores on the more discriminating scales, and provided estimates of predictive powers under a range of clinically realistic base rates. Multilevel likelihood ratios combine the information

about the diagnostic sensitivity and specificity of test scores in a given range, packaging the data in a way that facilitates using Bayes Theorem to estimate revised probabilities of diagnoses. We provide a clinical vignette in the Discussion to illustrate the potential clinical utility of these methods for decision making about individual cases.

Method

Participants

Youths aged 5 to 18 years were recruited for studies on childhood psychiatric disorders. The only eligibility requirements were that both the patient and their caregiver were able to speak English; however, participants were excluded if they suffered from a pervasive developmental disorder, or mental retardation.

The first sample ($N=1084$) was recruited from a psychiatric research center with a focus on bipolar disorders, and referrals of offspring from parents seen at an affiliated adult mood disorders clinic (Findling et al., 2005; Youngstrom et al., 2005). Families completed the semi-structured diagnostic interview after a phone screen determined potential eligibility for ongoing treatment studies (Findling et al., 2005; Youngstrom et al., 2005).

The second sample ($N=651$) was a consecutive case series recruited from an urban community mental health center that primarily served African-American families living in the inner-city region (Youngstrom et al., 2005). Table 1 reports descriptive statistics by sample.

Parents and youth in both samples were led through an informed consent process, after which they were asked to provide their consent and assent, respectively. Families were provided with compensation for their time. All measures included in the present study were collected at the baseline visit, consequently, there was no attrition.

Measures

Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS)—All participants and their parents were interviewed using the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Epidemiological version (K-SADS-E; Orvaschel, 1994), or the Present and Lifetime version (K-SADS-PL; Kaufman et al., 1997). The interviews were conducted by highly-trained research assistants. All diagnoses were reviewed by a licensed child psychologist and/or psychiatrist. Diagnoses were blind to scores on the behavior checklists; checklists and KSADS were gathered at the same visit.

Child Behavior Checklist (CBCL)—Parents completed the CBCL about their child (Achenbach, 1991a; Achenbach & Rescorla, 2001). The CBCL has 118 problem behavior items rated from 0 (*Not True (as far as you know)*) to 2 (*Very True or Often True*), items were scored according to standard practices (Drotar, Stein, & Perrin, 1995). Data collection used the 1991 version, switching to the 2001 version when it became available (Youngstrom et al., 2005). The majority of the items remained the same, particularly on the Internalizing and related scales. The present study focused on scales related to anxiety. Reliability was

acceptable in the present data: Internalizing, Cronbach's $\alpha=.88$; Anxious/Depressive, $\alpha=.80$; Withdrawn, $\alpha=.79$; Thought Problems, $\alpha=.77$; Attention Problems, $\alpha=.82$; Social Problems, $\alpha=.76$; Somatic Complaints, $\alpha=.75$; DSM Anxiety Problems, $\alpha=.67$; DSM Affective Problems, $\alpha=.73$.

Youth Self Report (YSR)—Youths aged 11 to 17 completed the YSR (Achenbach, 1991b; Achenbach & Rescorla, 2001). The YSR has nearly identical content to the CBCL, organized into similar scales. Again, data collection used the 1991 version until the 2001 version was available. Reliability was similarly acceptable for the scales used here: Internalizing, $\alpha=.90$; Anxious/Depressive, $\alpha=.80$; Withdrawn, $\alpha=.74$; Thought Problems, $\alpha=.79$; Attention Problems, $\alpha=.78$; Social Problems, $\alpha=.74$; Somatic Complaints, $\alpha=.78$; DSM Anxiety Problems, $\alpha=.66$; DSM Affective Problems, $\alpha=.80$.

Teacher Report Form (TRF)—Families also picked the teacher most familiar with the child and asked them to complete the Achenbach TRF (Achenbach, 1991c; Achenbach & Rescorla, 2001). The TRF has nearly identical items and scales to the CBCL. Reliability was similarly acceptable for the scales used here: Internalizing $\alpha=.93$, Anxious/Depressed $\alpha=.84$, Withdrawn $\alpha=.80$, Thought Problems $\alpha=.81$, Attention Problems $\alpha=.94$, Social Problems $\alpha=.81$, and Somatic Complaints $\alpha=.96$.

Procedure

In both samples, youths and their primary caregiver completed the K-SADS interview. The Longitudinal Evaluation of All Available Data (LEAD) standard of diagnosis was used to finalize all diagnoses in the study (Spitzer, 1983). The LEAD diagnoses integrated information collected through the K-SADS interview, family history, prior treatment history, and clinical judgment. Kappa was 0.91 for all diagnoses when LEAD diagnosis was compared to the K-SADS diagnosis (Youngstrom et al., 2005). Additionally, each caregiver completed a CBCL about their child, and youths aged 11 years and older completed the YSR. The teacher most familiar with the youth also completed packet of questionnaires including the Teacher Report Form (TRF) version of the Achenbach.

Analytic Plan

Chi-squared and *t*-tests compared the two samples in terms of demographic and clinical characteristics. Receiver operating characteristic (ROC) analyses (Kraemer, 1992; McFall & Treat, 1999; Youngstrom, in press) assessed the diagnostic efficiency of each of the CBCL, YSR, and TRF subscales, for determining diagnoses of any Anxiety Disorder, Generalized Anxiety Disorder, and Specific Phobia. Anxiety disorder diagnoses were included in all analyses regardless of comorbidity or referral question. We inspected score distributions and ROC curves for indications of “degenerate distributions,” where extreme scores on the index test might occur in cases without anxiety disorders (Youngstrom, in press; Zhou, Obuchowski, & McClish, 2002). Other anxiety disorders, such as OCD, were not analyzed separately due to low prevalence in the present samples.

Because the focus was on anxiety disorders, we omitted the Externalizing problems, Total problems, Aggressive Behavior and Delinquent Behavior (renamed Rule Breaking Behavior

on the 2001 versions), as well as DSM oriented scales focused on externalizing behavior problems. These scales were not significantly correlated with any anxiety disorder or with GAD (point biserial r values ranging from $-.08$ to $.05$).

Those scales performing better than chance ($AUC > .50$) were compared to evaluate which was the most discriminating measure for each anxiety diagnosis using the t -test for dependent AUCs (Hanley & McNeil, 1983). The AUCs for each scale were compared across the two samples, using the z -test of independent AUCs (Hanley & McNeil, 1983). If no significant differences were found, subsequent analyses combine the samples to provide smaller standard errors and more precise estimates. We organized analyses using the top-down framework for test interpretation (Sattler, 2002; Watkins, 2009; Youngstrom, 2008), giving priority to more global scores and simpler algorithms unless subscales or combinations of scales could demonstrate statistically significant incremental validity. For any test demonstrating statistically significant AUCs, the diagnostic likelihood ratio (DLR) was calculated, along with positive predictive value for each diagnosis from the Internalizing T -Score. Logistic regression analyses tested the incremental validity of combinations of scales.

Complete data were available within informant. We chose not to impute data for youth without YSR scores because the YSR was not intended for use in the younger age group, does not have normative data, and is only used “off label” if at all in this age range. We also decided not to impute scores for teachers missing the TRF because there were enough missing reports that imputation created large standard errors and did not improve power for results. Youth who completed the self report were older, more female, had more depression and less ADHD or ODD (consistent with all the main effects of age and referral pattern) than youth who did not complete the YSR; teacher report did not show evidence of any pattern of missing data.

Results

Table 1 reports the demographic and clinical characteristics of both samples. Participants in the community clinic were significantly younger by roughly a year on average. As anticipated based on the referral patterns, the academic clinic included a significantly larger percentage of white families, and the community clinic included significantly more black families. The academic clinic sample included significantly more major depressive disorder and dysthymia, as well as more bipolar spectrum disorders. The community clinic sample included significantly more anxiety disorders, oppositional defiant disorder, attention deficit hyperactivity disorder; youths in the community clinic also met criteria for more axis I diagnoses on average.

Diagnostic Efficiency

Anxiety disorders were present in 13% of the academic clinic sample ($n = 141$) and 26% of the community clinic sample ($n = 165$). However, only two specific anxiety disorders, generalized anxiety disorder and specific phobia, were sufficiently prevalent to have at least 20 cases occur in both settings, satisfying Kraemer’s (1992) rule of thumb for a minimally adequate sample size to estimate diagnostic efficiency parameters. None of the CBCL or

YSR scales discriminated specific phobia at better than chance levels (results available upon request from the authors). Similarly, none of the TRF scales discriminated any of the anxiety criteria at better than chance levels in either sample (results also available upon request from the authors). The CBCL and YSR Internalizing problems *T* scores discriminated cases with any anxiety disorder or with GAD from all other diagnoses in both samples; see Table 2 for discernment of any anxiety disorder versus all other cases, and Table 3 for results with GAD. Though the CBCL and YSR discriminated any anxiety or GAD from other diagnoses, the AUCs for these scales fell primarily under “low” or low-medium discriminatory ability according to Swets’ (1988) benchmarks. The Cohen’s *d* values for the same comparisons would conventionally be considered “medium” ($d \sim .5$) to “large” ($d \sim .8$), with estimates ranging from .46 to .91.

The clinical syndrome scales underlying the Internalizing Problems broadband – Anxious/Depressed, Withdrawn, and Somatic Complaints – also tended to be significant, but not better at discriminating than the other scale scores. The presence of any anxiety disorder also was associated with significant elevations on the Thought Problems, Attention Problems, and Social Problems clinical syndrome scales, but these were of significantly smaller magnitude than the AUCs observed for Internalizing and for the Anxious/Depressed scales. The DSM scales – Anxiety Problems and Affective Problems – performed similarly to the Internalizing and Anxious/Depressed scales, with AUCs ranging from .60 to .68 for Any Anxiety and from .59 to .70 for GAD.

Examination of the score distributions found some indication of “degenerate” distributions. In this context, “degenerate” refers to situations where high scores occur frequently in the comparison group, reducing the diagnostic specificity high scores. For example, many of the high scoring cases on Internalizing did not have anxiety disorders, but did have depression. Nonparametric ROC estimation makes few distributional assumptions; but when the comparison group has significantly larger variation in scores, or if there are outliers with high scores in the comparison group, then it will be impossible to achieve good discrimination between diagnostic groups in the high score range (Pepe, 2003; Youngstrom, in press; Zhou et al., 2002). In both samples and across all measures, cases with mood disorders also showed high scores on Internalizing and the other scales, with the means equal the means for the group with anxiety disorders but no comorbid mood. The non-anxiety group also had significantly larger variances and more cases with extreme high scores (*T* scores of 80+) than did the subgroup with anxiety diagnoses, reflecting the greater prevalence of mood disorders than anxiety disorders in both clinical settings (see Figure 3). Degeneracy does not invalidate the overall ROC analysis, but suggests that the performance of the test will be much more useful in some score ranges than others. Our analyses addressed the degeneracy by examining the likelihood ratios and pooling score intervals where the likelihood ratios did not rise steadily (Zhou et al., 2002).

Comparisons of the AUCs within each sample established that there were no significant differences in the discriminative validity of the CBCL versus YSR Internalizing scores (p values $> .05$), and both were superior to the TRF Internalizing ($p < .0005$) for both the any anxiety and the GAD criteria.

The *t*-test of dependent ROCs indicated that for GAD, the Anxious/Depressed score performed slightly better than the Internalizing score ($z=2.53, p=.011$). Additionally, the DSM Anxiety Problems scale, outperformed the Internalizing scale at identifying Any Anxiety ($z=3.19, p=.001$). For every other comparison, the Internalizing subscale performed as well or better than the other scales.

The diagnostic efficiency of the CBCL and YSR scales were not statistically different between boys and girls. Additionally, with the exception of the Anxious/Depressed CBCL scale, the scales performed equally well in the Academic and Community samples. The AUC for the Anxious/Depressed scale was higher in the Academic sample for both GAD ($z = 2.08, p = .038$) and any anxiety ($z = 3.03, p = .002$); however, this difference was not robust enough to survive post hoc correction for number of comparisons.

Incremental Validity

Logistic regression analyses tested whether combinations of scales significantly improved on the performance of the Internalizing scale in isolation. The combination of YSR and CBCL Internalizing scores predicted the “any anxiety” criterion, $X^2(2) = 43.54, p < .0005$. Both the YSR and the CBCL Internalizing scores made significant unique contributions, $B = .04, p < .0005$ for CBCL Internalizing, and $B = .03, p < .0005$ for YSR Internalizing. Saving the predicted values from the logistic regression and then using them in the ROC analysis yielded an AUC of .67 in the pooled sample of youths old enough to have YSR scores, not significantly different from the AUC of .64 for the CBCL or YSR scores in isolation. Simply averaging CBCL and YSR internalizing scores produced an AUC of .68, also not significantly different than either constituent score. This pattern of results indicates that the combination of CBCL and YSR scores leads to a statistically significant but clinically trivial change in diagnostic performance. A similar pattern of findings occurred when GAD served as the criterion: Both CBCL and YSR made statistically significant unique contributions, but the classification accuracy of the combination did not significantly improve on the performance of either in isolation.

Diagnostic likelihood ratios (DLRs) were calculated for score ranges corresponding to low, medium, and high risk for any anxiety disorder using the Internalizing scores from the CBCL and YSR. DLRs that are less than 1 are associated with test scores that indicate lower probability of disorder, whereas scores above 1 are associated with higher probabilities of the disorder. In our samples, low CBCL or YSR scores were associated with DLRs reducing the odds of an anxiety diagnosis, ranging from .10 to .25, where .1 might be considered clinically decisive that there is no anxiety disorder, and .20 would be considered moderately certain (Straus et al., 2011). High scores were less decisive in changing the odds of anxiety disorders. For individuals in the highest risk group, Internalizing scores >69 (CBCL) or >63 (YSR) resulted in a DLR of 1.5. See Table 4. The smaller DLRs for the high scores resulted from the degenerate distributions described above, where cases with mood disorders also scored high on the Internalizing and other scales, and occurred at similar rates as the cases with anxiety disorders in the higher score ranges (see Figure 3).

Discussion

The goal of the present study was to investigate the diagnostic efficiency of one of the most widely used cross-informant measures of psychopathology for the purpose of assessing potential anxiety disorders in children and adolescents. The study replicated prior investigations finding that the Achenbach CBCL and YSR showed discriminative validity for separating anxiety disorders from other cases seeking outpatient services. The present study extends prior work in several ways, including, (a) using the largest samples published yet with semi-structured diagnostic interviews as the criterion measure, (b) examining the generalizability of results from academic to community mental health settings with significantly different demographic and clinical characteristics, (c) directly comparing the performance of parent, youth, and teacher report on the instruments, (d) evaluating whether the integration of information from multiple informants provides significant incremental improvement with regard to identifying anxiety disorders, and (e) reporting the diagnostic likelihood ratios and other information to facilitate the direct application of test results to clinical decision-making about individual cases.

Results indicated that the CBCL and YSR scales discriminated cases with any anxiety disorder from other youths seeking services, whereas TRF scales did not perform at better than chance levels. Despite substantial differences in demography and referral patterns, these variables did not moderate the diagnostic validity of the CBCL and YSR scales, making it possible to pool samples and estimate a single set of diagnostic likelihood ratios that would generalize across both settings. Combining CBCL and YSR scores produced statistically significant improvement in prediction, although it is less clear that the incremental value has clinical significance.

Another key aspect of the present findings was that cases with mood disorders also produced high scores on the measures that putatively would be helpful in identifying anxiety disorders. The association between Internalizing scores and unipolar depression (Warnick et al., 2008) or bipolar disorder (Mick, Biederman, Pandina, & Faraone, 2003) is well known, and anxious and depressed symptoms load together on the Anxious/Depressed component in analyses of the Achenbach items (Achenbach & Rescorla, 2001; Lengua, Sadowski, Friedrich, & Fisher, 2001). The items on the Achenbach scales mostly reflect negative affect and general distress, which the tripartite model of depression and anxiety (Clark & Watson, 1991) has established are shared features, not specific to either set of diagnoses in youths (Chorpita, 2002; Lonigan, Phillips, & Hooe, 2003) as well as adults. This lack of specificity manifested as degenerate score distributions (Pepe, 2003), where cases with anxiety disorders scored high on scales, but so did cases with mood disorders (Figure 3). When item content focuses on negative affect, then high scores will be associated with both depression and anxiety (Ferdinand, 2008), and there is no score threshold that would clearly tease apart these two possibilities. Inconsistent findings in prior studies of the Achenbach scales as discriminating anxiety disorders may have been confounded by differences in the rate of mood disorder in the sample. Studies that systematically excluded mood disorder would increase the apparent diagnostic specificity of the scales by eliminating a major source of false positive scores (Youngstrom, Meyers, Youngstrom, Calabrese, & Findling, 2006; Zhou et al., 2002). Conversely, studies that included mood disorder would have more false

positives when hunting for anxiety disorder, but this would more accurately model how the scale would function in other settings with a similar mix of mood and anxiety disorders. Epidemiological studies indicate that anxiety disorders are more common than mood disorders before puberty, with the pattern reversing after age 10–12, and the overall lifetime rates of any anxiety disorder and any mood disorder both hovering around 9 to 14% in the general population of children and adolescents (Beesdo, Pine, Lieb, & Wittchen, 2010; Merikangas, He, Burstein, et al., 2010). Cases with mood disorder also may be somewhat more likely to seek services, suggesting that the ratio of mood disorders to anxiety disorders observed here may be fairly generalizable (Merikangas, He, Brody, et al., 2010).

Limitations

Limitations of the present study include relatively low rates of specific anxiety disorders, precluding the investigation of whether the Achenbach scales were particularly useful for differentiating panic or obsessive-compulsive disorders, for example. This concern is mitigated some by the observation that the more common diagnoses in epidemiological and general outpatient settings, such as GAD and phobia, were well-represented. The rate of “any anxiety” disorder was consistent with benchmarks from prior work, and provided good statistical power and precision for estimates of diagnostic performance (Kraemer, 1992). It also is important to note that the sample design included several features likely to attenuate diagnostic efficiency, but which enhance clinical generalizability, such as the limited exclusion criteria, high rates of comorbidity, and the inclusion of a large number of cases with diagnoses likely to generate false positive test results (Bossuyt et al., 2003). Another limitation is the fact that this study did not include Spanish-speaking participants. Though the sample was diverse from a racial and socioeconomic perspective, the exclusion of Spanish-speaking people limits the generalizability of the results to non-English speakers. Also, it is important to note that there are other well-established semi-structured interviews that have even more extensive validity data for anxiety disorders (e.g., ADIS, Silverman & Nelles, 1988). It is unclear whether using the ADIS instead of the KSADS would change results. Finally, the diagnostic efficiency of the Achenbach scales was limited by the low specificity of high scores to anxiety disorders. If other scales show greater diagnostic specificity to anxiety disorders, then high scores on them would do a better job of helping rule in an anxiety disorder, increasing the posterior probability (Straus et al., 2011).

Clinical implications

The results of the present study suggest that the CBCL does not provide sufficient information to aid in the diagnosis of *specific* anxiety disorders in clinical settings with a prevalence of anxiety disorders similar to the rates in our samples, 13% and 26%. However, the CBCL is often administered as part of a clinical intake procedure, and consequently results in no additional cost to clinic or family. So, though it might not be worthwhile to administer the CBCL or YSR for the sole purpose of identifying a specific anxiety disorder, these tools do provide information regarding the presence of *any* anxiety disorder, and given the low burden, may be useful to clinicians. This result is consistent with previous studies that have found the CBCL helpful at “ruling in or out” an anxiety disorder (Aschenbrand et al., 2005; Pauschardt et al., 2010). Presented with a new patient, clinicians typically generate between five and seven candidate diagnoses (Norman, 2009), and if the correct diagnosis is

part of the original hypotheses, the correct diagnosis is often chosen by the end of the evaluation process. The CBCL can be used to help develop a short list of diagnoses to consider. Internalizing and anxious symptoms are present in youth for other reasons besides anxiety; mood disorders are the most common, but adjustment problems, developmental disorders, and other factors could play a role; parsing out symptoms has important treatment implications. The CBCL can help with this, even in highly comorbid samples, like ours.

The information gleaned from the CBCL and YSR may be particularly helpful when combined with other information in an actuarial approach. An important strength of taking an actuarial approach to diagnostic decision making is that it allows for different sources of information to be incorporated in an objective manner. For example, taken alone, known risk factors for developing an anxiety disorder, including parent anxiety disorder, high behavioral inhibition, female gender, and high CBCL Internalizing scores, are not sufficient for diagnosis. But, when combined using the nomogram (Figure 1), these factors have predictive value that can help a clinician rule out an anxiety diagnosis or determine that a more specific anxiety assessment is necessary.

There is ample support for an evidence-based approach to diagnosis, but psychology and psychiatry have not made as much progress as other fields in utilizing “weak” signals to predict outcomes (Drake et al., 2001; Hoagwood, Burns, Kiser, Ringeisen, & Schoenwald, 2001; Hunsley & Mash, 2007). For example, the correlation between CBCL scales and any anxiety ($r=.22$) is similar to mammogram prediction of breast cancer two years later ($r=.27$), and better than IQ score predicting functional effectiveness across jobs ($r=.25$), and is equivalent to verbal GRE score predicting GPA ($r=.28$), yet these pieces of information are commonly used – along with other signals – to forecast health risk or academic and professional success (Neisser et al., 1996; Gottfredson, 1997; Lubinski, 2004).

In the case of childhood anxiety, prediction is important; some anxiety is normative among children, being able to identify cases for whom the anxiety is likely to subside over time, versus those for whom treatment is necessary, is another area in which the CBCL and YSR may be helpful. In evidence-based medicine, conditions may be categorized based on a similar idea, some require treatment, whereas others fall in “assess” or “wait and see” zones (Straus et al., 2011). A three-tiered assessment model has been developed, and successfully employed in the field of pediatric bipolar disorder. Youngstrom et al. (2013; Youngstrom, Jenkins, Jensen-Doss, & Youngstrom, 2012) proposed a stoplight system, whereby patients are categorized, based on risk, in order to determine next clinical actions: “Green” – minimal/no risk, “Yellow” – further assessment needed, and consider using broad-spectrum and low risk interventions, and “Red” – needs acute treatment. Rather than relying on an initial assessment and clinical intuition to make a final treatment decision, the EBA approach integrates assessment findings into a probability that then guides the next steps in terms of both assessment and treatment without unnecessary cost and burden to the clinic or the patient. A clinical vignette illustrates the application of these techniques and guiding principles.

Vignette

A 14-year-old girl is referred to the clinic by her teacher due to symptoms of withdrawal, poor attention, school attendance problems, and general worries. Her mother completed a CBCL and the patient completed the YSR. The CBCL Internalizing T score was 76, and her YSR T score was 70. In order to incorporate this information using the nomogram (See Figure 1), first select an appropriate pretest probability. Meehl (e.g., 1954) and others have recommended using the base rate of anxiety disorders, either in the community or in a clinical setting similar to this one, as the starting point for assessment. Next, determine the Diagnostic Likelihood Ratio (DLR) associated with a specific risk factor or with a test result, and plot it on the middle line of the nomogram. In this case, an average Internalizing T score of 70, based on her CBCL and YSR scores, is associated with a DLR of 2.67 (see Table 4, using the average of the two T scores). Then connect the dots between the pretest probability and the Internalizing DLR, and extend the line across the right-hand line to estimate the posterior probability (likelihood that the patient has an anxiety disorder, based on the base rate of anxiety disorders and her CBCL score), which is 34% in this case (see Figure 2). In order to add new information, such as family history of anxiety, put the posterior probability value as the new pretest probability, and repeat the steps, plotting the DLR associated with family history on the middle line. In this case, the patient's mother reports that she has been diagnosed with GAD and is currently being treated with psychotherapy and an SSRI. Anxiety disorders are heritable, with family members at a four-to-six times higher risk of developing an anxiety disorder (Smoller et al., 2008). For our patient, we will add a DLR of 5 to account for her family history of anxiety. Now, connect the dots between the initial posterior probability (34%) and extend through the DLR of 5 to determine the new posterior probability, 71%. The order in which risk factors are entered does not matter. In fact, if multiple distinct pieces of information are available at the same time, the associated DLR values can be multiplied together to estimate a single combined DLR, saving the need for several iterations through the nomogram process. The addition of the family history information raises the posterior probability to 71% (see Figure 2), falling in the "Yellow Zone" between the test and treat thresholds, indicating that more focused evaluation of anxiety disorders, along with low risk treatment, like psychotherapy, is an appropriate course forward (Youngstrom, 2013). For more information about the nomogram procedure, see Jenkins et al. (2011).

Conclusion

The CBCL and YSR are not the only questionnaires that assess for anxiety symptoms in young people; however, the ASEBA system is widely used and studied, making it an obvious starting point for the development of an evidence-based approach to diagnosing anxiety disorders in youth. However, future research should extend to other measures, particularly those that have a low burden to clinic and patient, in order to determine which is more diagnostically helpful. If another measure results in a bigger AUC, this would be a compelling reason to switch measures (McFall & Treat, 1999). Measures that focus on symptoms more specific to anxiety disorders, such as physiological hyperarousal and fear for panic disorder, or obsessions and compulsions, are likely to yield more diagnostic specificity, and thus may be more helpful in ruling in specific anxiety disorders. However,

this greater specificity needs to be set against the costs of longer assessment approaches and the low base rate of these conditions in many clinical settings. Rather than universal screening for rare conditions, results suggest that broad spectrum measures such as the Achenbach scales can help rule out anxiety disorders in a substantial portion of cases, while identifying a group of cases for additional evaluation with more specific and specialized methods.

The two samples in the present study represent broad demographic and clinical variation. Additionally, previous studies of the diagnostic ability of the CBCL have been inconsistent, in terms of the scales used and AUCs reported; replication can bolster the evidence in favor of the use of particular scales. It is also important to take into consideration the role of moderators. In the present study, we investigated clinical setting, gender, and age as potential moderators. The finding that these did not interact with diagnostic efficiency for detecting anxiety disorders may partly be due to the age and gender norms used to generate the *T* scores. Regardless, the lack of significant statistical moderation is good news for clinicians and families, as it indicates that the existing norms and research findings are likely to be applicable to a wide swath of youths (Jaeschke et al., 1994). In contrast, the high rate of mood disorders in the sample had a substantial effect on the diagnostic efficiency of the scales, indicating that this will be a key variable for clinicians to consider when applying research evidence to clinical cases.

It is valuable to take an “effectiveness,” rather than an “efficacy” approach to assessment research (Youngstrom, 2008). Even though the results are likely to be less impressive than what would be found in more finely-filtered samples, studies including a broad range of youth are more generalizable to clinical practice. Additionally, effectiveness-oriented research designs provide more accurate answers to the question of “will this help my patient”? (Jaeschke et al., 1994). Realistic expectations about the available information and its diagnostic validity will help clinicians approach cases with appropriate levels of caution and confidence, leading to better diagnoses and treatment.

Acknowledgments

We thank the families who participated in this research. This work was supported in part by NIH 5R01 MH066647 (PI: E. Youngstrom) and a center grant from the Stanley Medical Research Institute (PI: R. Findling). Dr. Youngstrom has received travel support from Bristol-Myers Squibb and consulted with Lundbeck. Dr. Findling receives or has received research support, acted as a consultant and/or served on a speaker's bureau for Alexza Pharmaceuticals, American Psychiatric Press, AstraZeneca, Bracket, Bristol-Myers Squibb, Clinsys, Cognition Group, Forest, GlaxoSmithKline, Guilford Press, Johns Hopkins University Press, Johnson & Johnson, KemPharm, Lilly, Lundbeck, Merck, NIH, Novartis, Noven, Otsuka, Oxford University Press, Pfizer, Physicians Postgraduate Press, Rhodes Pharmaceuticals, Roche, Sage, Seaside Pharmaceuticals, Shire, Stanley Medical Research Institute, Sunovion, Supernus Pharmaceuticals, Transcept Pharmaceuticals, Validus, and WebMD.

References

- Achenbach T. Manual for the child behavior checklist. 1991a
- Achenbach, T. Manual for the Youth Self-Report and 1991 profile. Dept. of Psychiatry, University of Vermont; 1991b.
- Achenbach, TM. Manual for the Teacher's Report Form and 1991 profile. Burlington: University of Vermont, Department of Psychiatry; 1991.
- Achenbach TM. Bibliography of published studies using ASEBA instruments. 2000

- Achenbach TM. Advancing assessment of children and adolescents: commentary on evidence-based assessment of child and adolescent disorders. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:541–547. [PubMed: 16026217]
- Achenbach, TM.; Rescorla, LA. *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont; 2001.
- Aschenbrand SG, Angelosante AG, Kendall PC. Discriminant validity and clinical utility of the CBCL with anxiety-disordered youth. *Journal of Clinical Child and Adolescent Psychology*. 2005a; 34:735–746. [PubMed: 16232070]
- Beesdo K, Pine DS, Lieb R, Wittchen HU. Incidence and risk patterns of anxiety and depressive disorders and categorization of generalized anxiety disorder. *Archives of General Psychiatry*. 2010; 67:47–57. [PubMed: 20048222]
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, de Vet HCW. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *British Medical Journal*. 2003; 326:41–44. [PubMed: 12511463]
- Chambless D, Ollendick T. Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*. 2001; 52:685–716.
- Chorpita BF. The tripartite model and dimensions of anxiety and depression: an examination of structure in a large school sample. *Journal of Abnormal Child Psychology*. 2002; 30:177–190. [PubMed: 12002397]
- Clark LA, Watson D. Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*. 1991; 100:316–336. [PubMed: 1918611]
- Cohen L, La Greca A, Blount R, Kazak A, Holmbeck G, Lemanek K. Introduction to Special Issue: Evidence-based Assessment in Pediatric Psychology. *Journal of Pediatric Psychology*. 2008; 33:911–915. [PubMed: 16675715]
- Dawes R, Faust D, Meehl P. Clinical versus actuarial judgment. *Science*. 1989; 243:1668–1674. [PubMed: 2648573]
- Drake RE, Goldman HH, Leff HS, Lehman AF, Dixon L, Mueser KT, Torrey WC. Implementing evidence-based practices in routine mental health service settings. *Psychiatric Services*. 2001; 52:179–182. [PubMed: 11157115]
- Drotar D, Stein REK, Perrin EC. Methodological issues in using the child behavior checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology*. 1995; 24(2):184–192.
- Ebesutani C, Bernstein A, Nakamura B, Chorpita B, Higa-McMillan C, Weisz J. Concurrent Validity of the Child Behavior Checklist DSM-Oriented Scales: Correspondence with DSM Diagnoses and Comparison to Syndrome Scales. *Journal of Psychopathology and Behavioral Assessment*. 2010; 32(3):373–384. [PubMed: 20700377]
- Ferdinand RF. Validity of the CBCL/YSR DSM-IV scales Anxiety Problems and Affective Problems. *Journal of Anxiety Disorders*. 2008; 22:126–134. [PubMed: 17321103]
- Findling RL, Youngstrom EA, McNamara NK, Stansbrey RJ, Demeter CA, Bedoya D, Calabrese JR. Early symptoms of mania and the role of parental risk. *Bipolar Disorders*. 2005; 7:623–634. [PubMed: 16403188]
- Frick PJ, Silverthorn P, Evans C. Assessment of childhood anxiety using structured interviews: Patterns of agreement among informants and association with maternal anxiety. *Psychological Assessment*. 1994; 6:372–379.
- Garb, HN. *Studying the clinician: Judgment Research and Psychological Assessment*. Washington, DC: American Psychological Association; 1998.
- Gottfredson LS. Why g matters: The complexity of everyday life. *Intelligence*. 1997; 24:79–132.
- Grove, W. The reliability of psychiatric diagnosis. In: Last, C.; Hersen, M., editors. *Issues in diagnostic research*. New York, NY: Plenum Press; 1987. p. 99-119.
- 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]

- Hoagwood K, Burns BJ, Kiser L, Ringeisen H, Schoenwald SK. Evidence-based practice in child and adolescent mental health services. *Psychiatric Services*. 2001; 52:1179–1189. [PubMed: 11533391]
- Hunsley J, Mash EJ. Evidence-Based Assessment. *Annual Review of Clinical Psychology*. 2007; 3:29–51. doi:
- Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test: B: What are the results and will they help me in caring for my patients? *JAMA*. 1994; 271:703–707. [PubMed: 8309035]
- Jenkins M, Youngstrom E, Washburn J, Youngstrom J. Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology: Research and Practice*. 2011; 42:121.
- 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. [PubMed: 9204677]
- Kessler R, Berglund P, Demler O, Jin R, Merikangas K, Walters E. Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 2005; 62:593–602. [PubMed: 15939837]
- Kraemer, HC. *Evaluating medical tests*. Newbury Park, CA: Sage; 1992.
- Lengua LJ, Sadowski CA, Friedrich WN, Fisher J. Rationally and empirically derived dimensions of children's symptomatology: Expert ratings and confirmatory factor analyses of the CBCL. *Journal of Consulting and Clinical Psychology*. 2001; 69:683–698. [PubMed: 11550734]
- 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.
- Lonigan CJ, Phillips B, Hooe E. Relations of positive and negative affectivity to anxiety and depression in children: Evidence from a latent variable longitudinal study. *Journal of Consulting and Clinical Psychology*. 2003; 71:465–481. [PubMed: 12795571]
- Lubinski D. Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "General intelligence," objectively determined and measured". *Journal of Personality and Social Psychology*. 2004; 86:96–111. [PubMed: 14717630]
- McFall RM, Treat TA. Quantifying the information value of clinical assessment with signal detection theory. *Annual Review of Psychology*. 1999; 50:215–241.
- Meehl, PE. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press; 1954.
- Meehl PE, Rosen A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*. 1955; 55:194–216. [PubMed: 14371890]
- Merikangas K, He J, Brody D, Fisher P, Bourdon K, Koretz D. Prevalence and Treatment of Mental Disorders Among US Children in the 2001–2004 NHANES. *Pediatrics*. 2010; 125:75–81. [PubMed: 20008426]
- Merikangas K, He J, Burstein M, Swanson S, Avenevoli S, Cui L, Swendsen J. Lifetime Prevalence of Mental Disorders in U.S. Adolescents: Results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*. 2010; 49:980–989. [PubMed: 20855043]
- Merikangas KR, He JP, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics*. 2010; 125:75–81. [PubMed: 20008426]
- Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, Swendsen J. Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*. 2010; 49:980–989. doi: S0890-8567(10)00476-4 [pii]. [PubMed: 20855043]

- Mick E, Biederman J, Pandina G, Faraone SV. A preliminary meta-analysis of the child behavior checklist in pediatric bipolar disorder. *Biological Psychiatry*. 2003; 53:1021–1027. [PubMed: 12788247]
- Neisser U, Boodoo G, Bouchard TJ Jr, Boykin AW, Brody N, Ceci SJ, Urbina S. Intelligence: Knowns and unknowns. *American Psychologist*. 1996; 51:77–101.
- Norman G. Dual processing and diagnostic errors. *Advances in Health Science Education Theory & Practice*. 2009; 14:37–49.
- Orvaschel, H. Schedule for affective Disorders and schizophrenia for School-Age Children-Epidemiologic Version (Vol. Fifth revision). Fort Lauderdale, FL: Nova Southeastern University; 1994.
- Pauschardt J, Remschmidt H, Mattejat F. Assessing child and adolescent anxiety in psychiatric samples with the Child Behavior Checklist. *Journal of Anxiety Disorders*. 2010; 24:461–467. [PubMed: 20362414]
- Pepe, MS. The statistical evaluation of medical tests for classification and prediction. New York, NY: Wiley; 2003.
- Rettew D, Lynch AD, Achenbach T, Dumenci L, Ivanova M. 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. [PubMed: 19701924]
- Sakolsky D, Birmaher B. Pediatric anxiety disorders: management in primary care. *Current Opinion in Pediatrics*. 2008; 20:538–543. [PubMed: 18781116]
- Sattler, JM. >Assessment of children: Behavioral and Clinical Applications. 4th ed.. La Mesa, CA: Author; 2002.
- Silverman WK, Nelles WB. The Anxiety Disorders Interview Schedule for Children. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1988; 27(6):772–778. [PubMed: 3198566]
- Smoller JW, Gardner-Schuster E, Misiaszek M. Genetics of anxiety: Would the genome recognize the DSM? *Depression and Anxiety*. 2008; 25(4):368–377. [PubMed: 18412063]
- Spitzer RL. Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry*. 1983; 24:399–411. [PubMed: 6354575]
- Straus, S.; Tetroe, J.; Graham, ID. Knowledge translation in health care: moving from evidence to practice. London: BMJ Books; 2011.
- Straus, SE.; Glasziou, P.; Richardson, WS.; Haynes, RB. Evidence-based medicine: How to practice and teach EBM. 4th ed.. New York, NY: Churchill Livingstone; 2011.
- Straus SE, McAlister FA. Evidence-based medicine: a commentary on common criticisms. *Canadian Medical Association Journal*. 2000; 163:837–841. [PubMed: 11033714]
- Swets J. Measuring the accuracy of diagnostic systems. *Science*. 1988; 240:1285–1293. [PubMed: 3287615]
- Warnick EM, Bracken MB, Kasl S. Screening Efficiency of the Child Behavior Checklist and Strengths and Difficulties Questionnaire: A Systematic Review. *Child and Adolescent Mental Health*. 2008; 13:140–147.
- Watkins, MW. Errors in diagnostic decision making and clinical judgment. In: Reynolds, CR.; Gutkin, TB., editors. *Handbook of School Psychology*. Wiley; 2009. p. 210-229.
- Youngstrom, EA. Evidence-based strategies for the assessment of developmental psychopathology: measuring prediction, prescription, and process. In: Miklowitz, D.; Craighead, W.; Craighead, L., editors. *Developmental Psychopathology*. New York, NY: Wiley; 2008. p. 34
- Youngstrom EA. Future directions in psychological assessment: Combining Evidence-Based Medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child & Adolescent Psychology*. 2013; 42:139–159. [PubMed: 23153181]
- Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*. (in press).
- 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. [PubMed: 15213586]

Youngstrom EA, Jenkins MM, Jensen-Doss A, Youngstrom JK. Evidence-based assessment strategies for pediatric bipolar disorder. *Israel Journal of Psychiatry & Related Sciences*. 2012; 49:15–27. [PubMed: 22652926]

Youngstrom EA, Meyers O, Demeter C, Youngstrom J, Morello L, Piiparinen R, Findling R. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders*. 2005; 7:507–517. [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. [PubMed: 17056395]

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

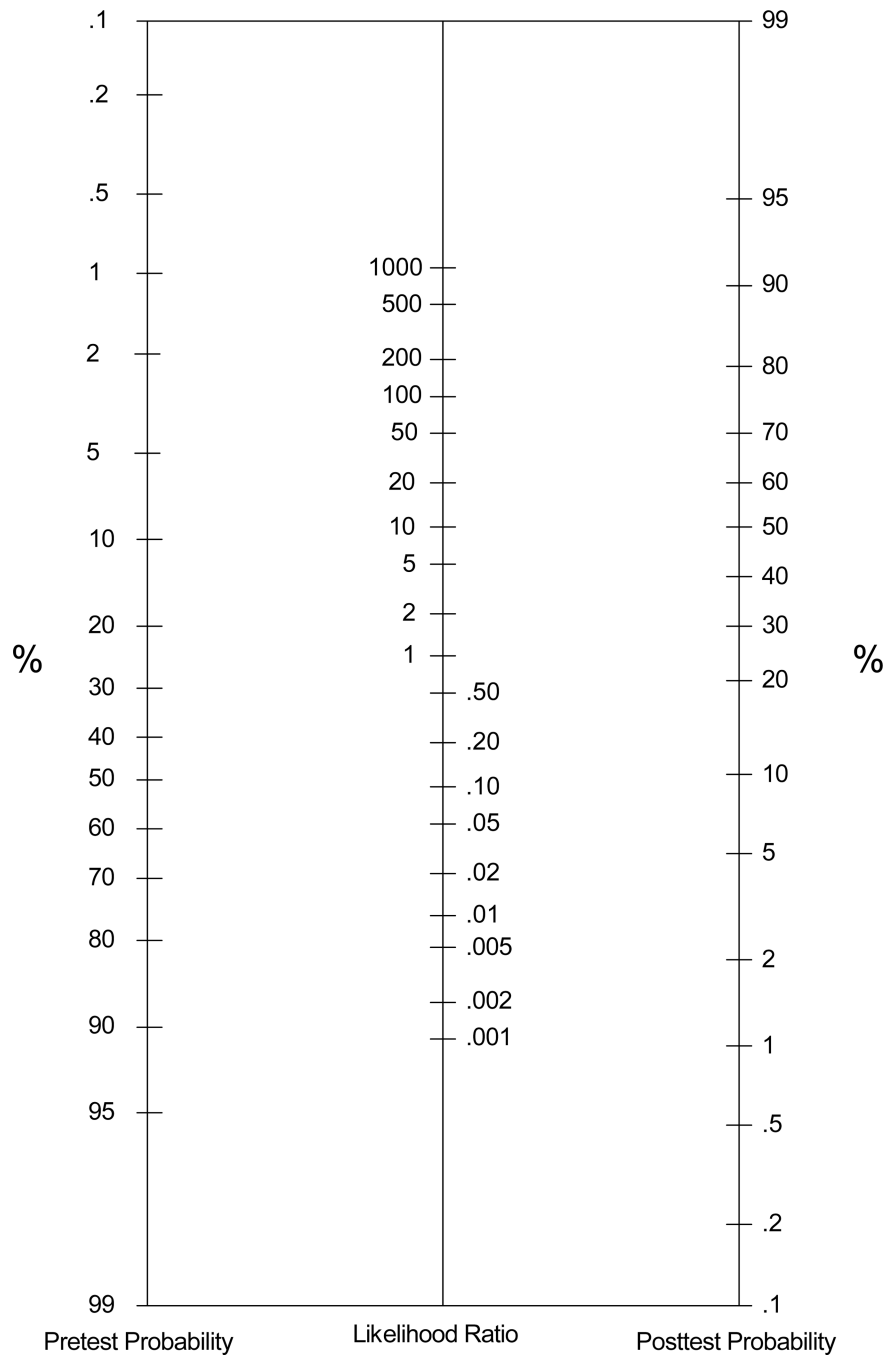


Figure 1. A probability nomogram for combining diagnostic likelihood ratios with other information about an individual case.

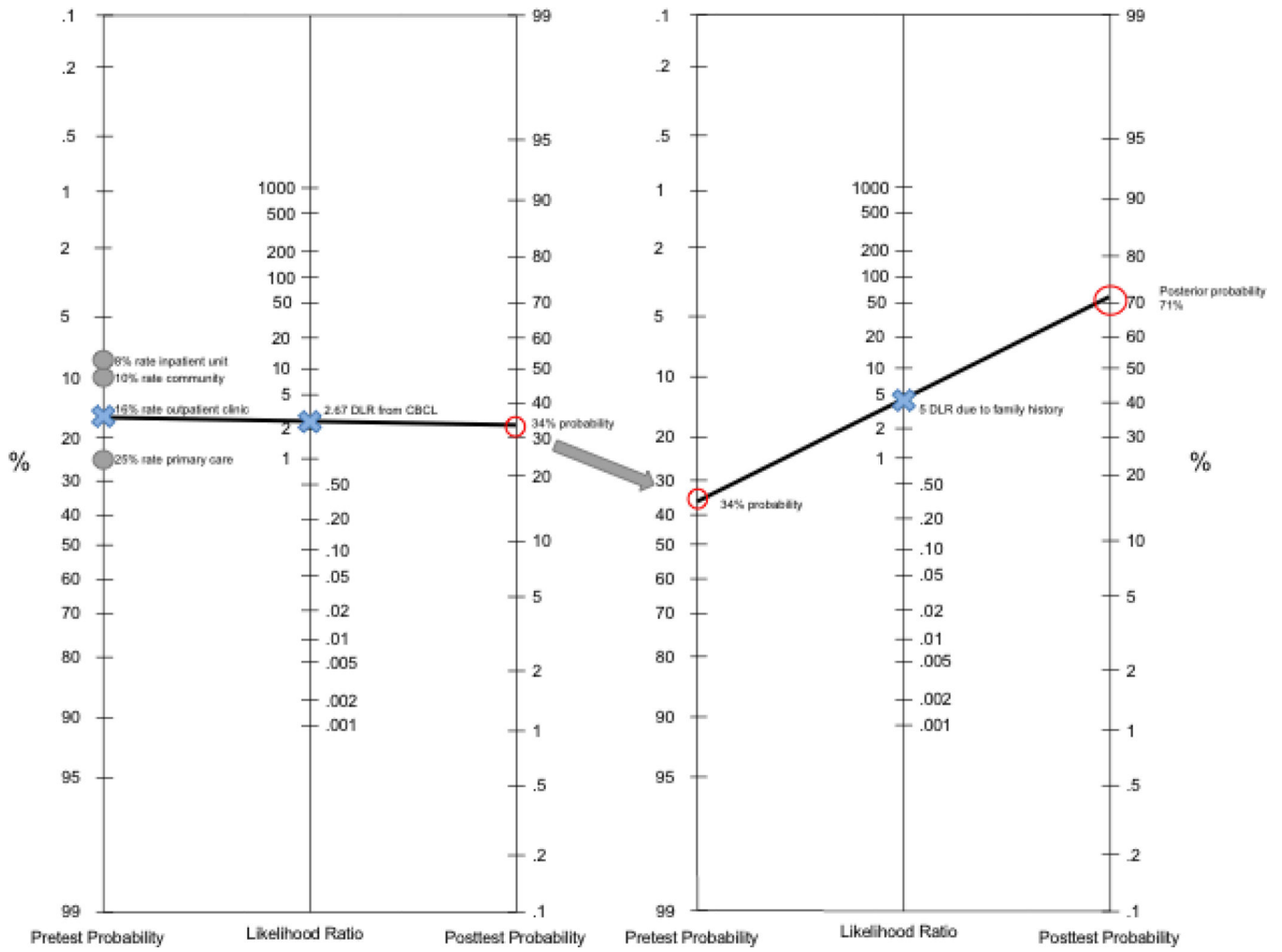


Figure 2.
Completed nomogram example from vignette.

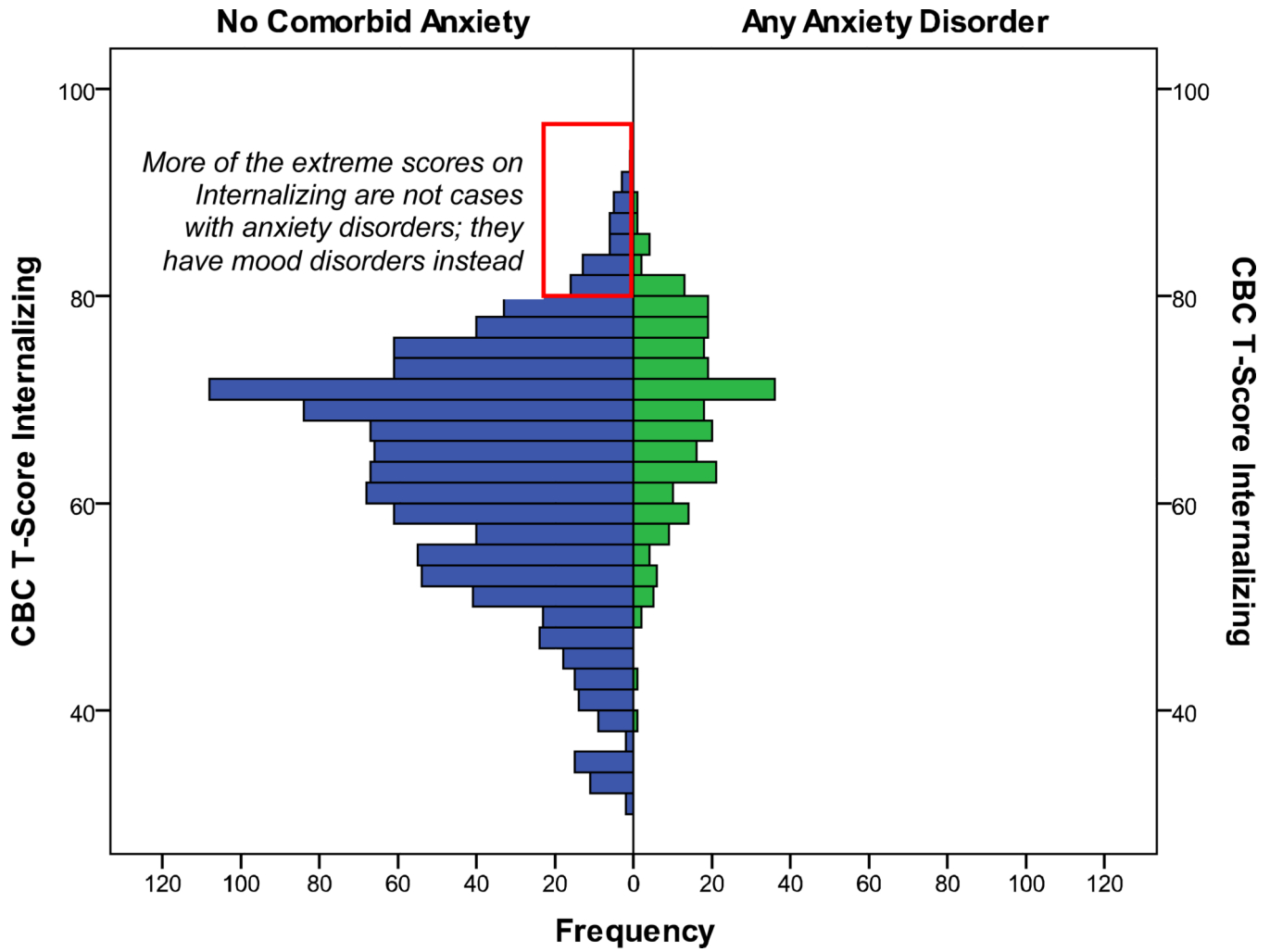


Figure 3. Back to back histogram of CBCL Internalizing score distributions for cases with any anxiety disorder diagnosis versus all other cases. Note that the distribution of scores for cases with an anxiety disorder tends to be shifted higher than the bulk of the distribution for cases with no comorbid anxiety, consistent with Internalizing scores being valid for discriminating anxiety disorders. However, the cases with the highest Internalizing scores do not have an anxiety disorder, indicating that the distribution is “degenerate” (Zhou et al., 2002).

Table 1

Demographic and Clinical Information Presented Separately by Clinical Setting

	Academic Clinic	Community Clinic
<i>N</i>	1084	651
Youth Age in Years (<i>SD</i>)	11.4 (3.4)**	10.6 (3.4)
Youth Gender (Male %)	62%	60%
Race		
White	79%***	7%
Black	14%	85%***
Hispanic	3%	2%
Other	4%	6%
Prevalence rate of any anxiety disorder	13%	26%***
Generalized anxiety disorder (GAD)	4%	4%
Specific phobia	2%	5%*
Separation Anxiety	1%	4%
Other diagnoses		
Major depressive disorder (MDD) and dysthymia	16%	29%***
Oppositional defiant disorder (ODD)	31%	38%**
Attention deficit/hyperactivity disorder (ADHD)	58%	65%*
Conduct disorder (CD)	10%	13%*
Bipolar spectrum disorders	48%***	14%
Number Axis I diagnoses (<i>SD</i>)	2.1 (1.3)	2.7 (1.4)***

Note.

* $p < .05$,** $p < .005$,*** $p < .0005$, two-tailed;

based on *t*-test for continuous variables (age, number of diagnoses) and chi-squared for categorical variables (gender, race, diagnostic group) comparing the academic (Findling et al., 2005; Youngstrom et al., 2005) to the community clinic samples (Youngstrom et al., 2005).

Table 2
 Diagnostic Efficiency of the Achenbach Scales at Discriminating Any Anxiety Diagnosis from All Other Diagnoses, Pooling Results from Both Samples (N = 1735)

Scale	Areas Under the ROC Curve				
	Academic Clinic	Community Clinic	Pooled Data	Standard Error	95% Confidence Interval
<i>CBCL</i>					
Internalizing	.69***	.63***	.64***	.02	.61 to .68
Anxious/Depressive	.74***, <i>a</i>	.64***	.66***	.02	.63 to .70
Withdrawn	.59**	.55*	.57***	.02	.54 to .61
Thought Problems	.63***	.57*	.61***	.02	.58 to .65
Attention Problems	.55*	.53	.55*	.02	.51 to .59
Social Problems	.58*	.58**	.59***	.02	.55 to .62
Somatic Complaints	.59**	.63***	.59***	.02	.56 to .63
<hr/>					
DSM Anxiety Problems ^b	–	.68***	–	.02	.64 to .73
<hr/>					
DSM Affective Problems ^b	–	.60***	–	.02	.55 to .65
<hr/>					
<i>YSR</i>					
Internalizing	.64***	.66***	.64***	.02	.59 to .69
Anxious/Depressive	.65**	.64***	.62***	.02	.58 to .67
Withdrawn	.65***	.64***	.65***	.02	.61 to .70
Thought Problems	.60	.63***	.62***	.02	.57 to .67
Attention Problems	.57	.64***	.59**	.02	.54 to .64
Social Problems	.63**	.58*	.61***	.02	.56 to .66
Somatic Complaints	.57	.64***	.60***	.03	.55 to .65
DSM Anxiety Problems ^b	–	.60**	–	.03	.55 to .66

Areas Under the ROC Curve

Scale	Academic Clinic	Community Clinic	Pooled Data	Standard Error	95% Confidence Interval
DSM Affective Problems ^b	–	.63***	–	.03	.57 to .68

Note.

* $p < .05$,

** $p < .005$,

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

Findling et al. (2005) used the 1991 version of the Achenbach scales, which did not include the DSM oriented subscales; Youngstrom et al. (2005) used the 2001 version.

^a Academic Clinic AUC significantly greater than Community Clinic AUC, $z = 3.03$, $p = .002$. Note that this difference would not survive post hoc correction for number of comparisons.

^b The DSM-Oriented scales were only available in the later protocol (Youngstrom et al., 2005), which used the 2001 version of the Achenbach instruments.

Table 3

Diagnostic efficiency of the Achenbach Scales at Discriminating Generalized Anxiety Disorder from All Other Diagnoses, Pooling Results from Both Samples (N = 1735)

Areas under the ROC Curve						
Scale	Academic Clinic	Community Clinic	Pooled Data	Standard Error	95% Confidence Interval	
<i>CBCL</i>						
Internalizing	.72***	.64*	.69***	.04	.62 to .76	
Anxious/Depressive	.80*** ^a	.64***	.74***	.04	.67 to .81	
Withdrawn	.62*	.57	.60*	.04	.52 to .67	
Thought Problems	.60	.57	.58	.04	.50 to .66	
Attention Problems	.50	.50	.49	.04	.40 to .58	
Social Problems	.56	.55	.55	.04	.47 to .63	
Somatic Complaints	.58	.68*	.62**	.04	.54 to .70	
DSM Anxiety Problems ^b	–	.70***	–	.05	.60 to .80	
DSM Affective Problems ^b	–	.59	–	.05	.50 to .69	
<i>YSR</i>						
Internalizing	.70*	.63	.67**	.05	.56 to .78	
Anxious/Depressive	.73***	.63*	.69***	.05	.59 to .79	
Withdrawn	.63*	.63	.62*	.05	.52 to .73	
Thought Problems	.51	.54	.53	.05	.43 to .63	
Attention Problems	.57	.66*	.61*	.05	.50 to .71	
Social Problems	.59	.54	.57	.05	.47 to .67	
Somatic Complaints	.64	.58	.61*	.05	.51 to .71	
DSM Anxiety Problems	–	.52	–	.06	.49 to .73	
DSM Affective Problems	–	.61	–	.06	.49 to .73	

Note.

* $p < .05$,

** $p < .005$,

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

Findling et al. (2005) used the 1991 version of the Achenbach scales, which did not include the DSM oriented subscales; Youngstrom et al. (2005) used the 2001 version.

^a Academic Clinic AUC significantly greater than Community Clinic AUC, $z = 2.08$, $p = .038$. Note that this difference would not survive post hoc correction for number of comparisons.

^b The DSM-Oriented scales were only available in the later protocol (Youngstrom et al., 2005), which used the 2001 version of the Achenbach instruments.

Table 4
 Diagnostic Likelihood Ratios (DLRs) Predicting Any Anxiety Disorder Diagnosis in the Pooled Sample (N = 1735)

Measure		Score Range				
		Low	Mod. Low	Neutral	Mod. High	High
CBCL Internalizing <i>T</i>	Score	<50	50 to 54	55 to 70	71 to 77	78+
	DLR	.13	.47	.98	1.51	2.03
YSR Internalizing <i>T</i>	Score	<42	42 to 50	51 to 65	66 to 72	73+
	DLR	.26	.57	.98	1.64	2.35
Average Internalizing	Score	<49	49 to 52	53 to 66	67 to 72	73+
	DLR	.10	.54	.88	1.61	2.67

Note.

There was no significant difference in the accuracy of the CBCL internalizing in the 5 to 10-year-old age group, AUC=.66, versus the 11 to 18-year-olds, AUC=.63, $z=0.78$, $p=.437$. Therefore we present one set of likelihood ratios across all ages from 5 to 18 years. Note that *T* scores already standardize scores based on age and gender norms; the prevalence rate of combined anxiety disorders was 19%.