



Published in final edited form as:

J Clin Child Adolesc Psychol. 2018 ; 47(SUP1): S161–S175. doi:10.1080/15374416.2016.1188698.

Diagnostic Efficiency of Caregiver Report on the SCARED for Identifying Youth Anxiety Disorders in Outpatient Settings

Anna Van Meter¹, Dokyoung S. You², Tate Halverson³, Eric A. Youngstrom³, Boris Birmaher⁴, Mary A. Fristad⁵, Robert A. Kowatch⁶, Amy Storfer-Isser⁷, Sarah M. Horwitz⁸, Thomas W. Frazier⁹, L. Eugene Arnold⁵, Robert L. Findling¹⁰, and the LAMS Group

¹Ferkauf Graduate School, Yeshiva University, Bronx, NY

²Texas A&M University, College Station, TX 77843

³University of North Carolina, Chapel Hill, NC 27599

⁴University of Pittsburgh School of Medicine, Pittsburgh, PA

⁵The Ohio State University, Columbus, OH 43210-1250

⁶The Ohio State University Wexner Medical Center, Columbus, OH

⁷Statistical Research Consultants, LLC, Schaumburg, IL

⁸New York University School of Medicine, New York, NY 10016

⁹Center for Pediatric Behavioral Health, Cleveland Clinic, Cleveland, OH 44195

¹⁰Johns Hopkins University/Kennedy Krieger Institute, Baltimore, MD 21205

Abstract

Objective—This study investigated the diagnostic and clinical utility of the parent-rated Screen for Child Anxiety Related Emotional Disorders (SCARED-P) for detecting youth anxiety disorders.

Method—Youth aged 6 to 12.9 were recruited from 9 outpatient mental health clinics ($N=707$). Consensus diagnoses were based on semi-structured interviews (KSADS) with youth and caregivers; 31% were diagnosed with at least one anxiety disorder. Caregivers completed the SCARED-P to describe youth anxiety levels. SCARED-P scores were not considered during the consensus diagnoses. Areas under the curve (AUCs) from receiver operating characteristic analyses and diagnostic likelihood ratios (DLRs) quantified performance of the SCARED-P total score and subscale scores (generalized anxiety disorder and separation anxiety disorder).

Results—SCARED-P total scores had variable efficiency (AUCs=.69-.88) and GAD and separation anxiety subscale scores were excellent (AUCs=.86-.89) for identifying specific anxiety disorders. Optimal subscale cutoff scores were computed to help rule-in (DLRs=2.7-5.4) or rule-out (DLRs<1.0) anxiety disorders among youth.

Conclusion—Results suggest that the GAD and separation anxiety SCARED-P subscales accurately identify their respective matched diagnoses. DLRs may aid clinicians in screening for youth anxiety disorders and improve accuracy of diagnosis.

Keywords

anxiety disorders; diagnostic accuracy; ROC; evidence based assessment

Anxiety disorders are common among youth, with lifetime prevalence estimated at 9-20% (Aschenbrand, Angelosante, & Kendall, 2005; Kessler et al., 2005; Merikangas et al., 2010; Sakolsky & Birmaher, 2008). They are associated with a number of negative outcomes for youth, both in the short- and long-term. In addition to increasing risk for internalizing disorders (anxiety, mood), and substance abuse disorders later in life (Beesdo-Baum et al., 2012; Grant et al., 2004; Shankman et al., 2009), anxiety may also interfere with an individual's ability to form relationships (Nelson, Rubin, & Fox, 2005), to pursue academic and professional goals, and to have a good quality of life. Though the burden of anxiety disorders is high (Baxter, Vos, Scott, Ferrari, & Whiteford, 2014), treatment is often successful (Hunsley, Elliott, & Therrien, 2014; Seligman, 2011) and can help to reduce the sequelae for both the individual and society.

Providing access to evidence-based treatment for a young person with an anxiety disorder has the potential to make a substantial difference in his/her future. However, accurate diagnosis of internalizing disorders in young people can be challenging. As with many other disorders, internalizing disorders are unlikely to fit a “prototypical” presentation and may manifest with a variety of symptoms and comorbid disorders (Kendall et al., 2010). Anxiety disorders, in particular, can be associated with agitation, irritability, and oppositionality, in addition to withdrawal, fear, and worry. Furthermore, some anxiety is developmentally appropriate and expected, so clinically significant anxiety can be difficult to distinguish (Sakolsky & Birmaher, 2008). Consequently, anxiety disorders in youth are often undiagnosed or misdiagnosed (Tomb & Hunter, 2004), especially if the child has a prominent externalizing disorder that preempts adult attention. In fact, youth with internalizing disorders are less likely to be brought for mental health services than their peers with externalizing disorders (Lynch & Clarke, 2006). Therefore, being able to estimate anxiety disorder risk using brief assessment tools that could be used in school or primary medical care settings may be particularly important to assure that those who could benefit from help receive it.

Many diagnostic assessments make use of screening tools to determine whether an individual is experiencing symptoms consistent with a certain disorder or class of disorders. However, information about *how* an individual's score influences his/her risk for a disorder (i.e., by how much does a score of 5 versus a score of 8 change the likelihood that a disorder is present?) is not often clear. Though an evidence-based approach will improve accuracy of the estimated likelihood of a diagnosis, information needed to take advantage of this approach can be difficult to find. Providing likelihood ratios for commonly-used clinical tools and demonstrating how to integrate them into an evidence-based approach to assessment may help to improve the assessment and diagnosis of childhood disorders, and

facilitate successful treatment (Christon, McLeod, & Jensen-Doss, 2014; Youngstrom, Choukas-Bradley, Calhoun, & Jensen-Doss, 2014). One effective, relatively simple way of incorporating evidence-based practices into assessment is to use a nomogram (Figure 1) to determine diagnostic probabilities (Straus, Glasziou, Richardson, & Haynes, 2011). The nomogram works by integrating likelihood ratios (based on known risk factors) into a posterior probability that a person meets criteria for a given disorder. This type of actuarial approach has repeatedly demonstrated superiority to clinical judgment (Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hamilton, 2001; Jenkins, Youngstrom, Washburn, & Youngstrom, 2011). Clinicians tend to make errors when estimating an individual's risk (Norman & Eva, 2010). For example, research has shown that, on average, a child's risk of having an anxiety disorder is 4 to 6 times higher if they have a parent with an anxiety disorder (Hettema, Neale, & Kendler, 2001; Rapee, 2012). For the average child, this would mean a probability of approximately 33% (9% base prevalence of anxiety among youth corresponds to odds of .99; quintupling the odds based on family history raises the odds to .49, corresponding to a probability of .33), but a clinician susceptible to confirmation bias might interpret the presence of parental anxiety as sufficient confirmation that the child has the same disorder. Confirmation bias is a major problem in clinical decision-making, and its effects with regard to family history were shown in a study of clinicians asked to estimate the likelihood that a youth had mood disorder, given a family history of bipolar disorder, in addition to symptoms and other risk factors (Jenkins et al., 2011). Use of a nomogram can help to effectively eliminate the influence of cognitive biases by providing probabilities based on data rather than interpretation. Additionally, the nomogram can incorporate multiple sources of data, which is advantageous in clinical settings where there is likely to be more than one measure of risk.

Among the tools available to assess anxiety in youth is the Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al., 1999). The SCARED is easy to complete and has been used successfully to screen for anxiety in a number of settings, including schools and primary care settings (Hale, Raaijmakers, Muris, & Meeus, 2005; Wren, Bridge, & Birmaher, 2004). There are both parent and child reports, with a five factor structure that corresponds to the following anxiety domains: panic/somatic, generalized anxiety, separation anxiety, social phobia, and school phobia. Importantly, the SCARED has demonstrated validity as an anxiety screening tool for youth from a variety of cultures and ethnicities (Canals, Hernández-Martínez, Cosi, & Domènech, 2012; Cosi, Canals, Hernández-Martínez, & Vigil-Colet, 2010; DeSousa, Salum, Isolan, & Manfro, 2013; Gonzalez, Weersing, Warnick, Scahill, & Woolston, 2012; Hale, Crocetti, Raaijmakers, & Meeus, 2011; Hariz, 2013). Additionally, the SCARED is both free and relatively brief, increasing its appeal for clinicians who are often constrained by limited time and resources.

Though several studies have investigated psychometric properties of the parent-rated SCARED (SCARED-P), relatively few studies have evaluated how a score on the SCARED-P would correspond with an anxiety disorder diagnosis as determined by a gold standard assessment approach. Studies that have evaluated sensitivity and specificity of the SCARED-P have found somewhat mixed results and have recommended different cut-off scores for identifying youth likely to meet criteria for an anxiety disorder. In an epidemiological study of Spanish youth, Cosi and colleagues (2010) found that parent report on the SCARED

subcales had small, but significant, correlations with anxiety symptoms measured by the M.I.N.I. Kid, a structured interview for youth. In a similar epidemiological study of Spanish youth, a score of 15-17 was recommended, corresponding to sensitivity of 63% to 66% and specificity of 61% to 70% for diagnoses made with the M.I.N.I (Canals et al., 2012). In contrast, in a study of Lebanese children at an outpatient mental health clinic, a cut-off score of 24 was recommended with sensitivity of 67% and specificity of 55% (Hariz, 2013), consistent with a study of treatment-seeking American adolescents that found an area under the curve (AUC) of .67 for the SCARED-P against diagnoses made with the K-SADS-P or a symptom checklist (Monga et al., 2000). Similarly, the parent-rated SCARED total score, somatic score, and separation anxiety score were able to distinguish youth with anxiety disorders diagnosed by a DSM-IV symptom checklist interview from youth without anxiety disorders ($ps < .005$) who were all receiving services at a mood and anxiety disorders outpatient clinic (Birmaher et al., 1999). Finally, in a study of youth (ages 8-15) presenting for primary care services, the SCARED-P total score showed fair diagnostic efficiency for identifying youth diagnosed with an anxiety disorder using the KSADS-PL (Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version) with an AUC of .79 (Gardner, Lucas, Kolko, & Campo, 2007).

Though results of these studies suggest that there could be clinical utility in using the SCARED-P as part of an assessment for a possible anxiety diagnosis, the relation between an individual's score on the SCARED-P and the likelihood that s/he has an anxiety disorder has not been fully explored. Furthermore, because many of the research studies using the SCARED to screen for anxiety disorders have focused on community or primary care populations, less is known about the SCARED-P's efficiency at identifying anxiety disorders among youth who are seeking services for mental health concerns. Outpatient mental health clinics are more likely than schools or primary care settings to have high rates of conditions likely to generate false positive scores on a scale. For example, mood disorders frequently generate high scores on measures of anxiety, even in the absence of a comorbid anxiety disorder (Clark & Watson, 1991), due to the fact that mood and anxiety disorders are both associated with high levels of negative affect, even in the absence of clinical symptoms of anxiety (e.g., Austin, Chorpita, & Chorpita, 2004, in child samples; Phillips, Lonigan, Driscoll, & Hooe, 2002). Testing an anxiety scale in a psychiatric clinical sample is likely to yield much more conservative estimates about its diagnostic specificity and lower overall estimates of diagnostic efficiency; but these estimates are also more likely to generalize to other clinical settings (Youngstrom, Meyers, Youngstrom, Calabrese, & Findling, 2006; Zhou, Obuchowski, & McClish, 2002). Considering that the SCARED-P is one of the most widely used rating scales for childhood anxiety, greater elaboration of its discriminative validity is important.

The present study was designed to determine how the SCARED-P performed in a sample of children (ages 6-12.9) seeking outpatient mental health services. This sample is significantly younger, on average, than previous outpatient mental health samples that have been used to evaluate the SCARED. Additionally, most youth in the sample were recruited because they had elevated symptoms of mania (though less than 23% met criteria for a bipolar spectrum disorder). Although this may limit the generalizability of the results, we believe that due to the high degree of overlap in symptoms between mania and other childhood disorders

(Kowatch, Youngstrom, Danielyan, & Findling, 2005) and the significant rate of comorbidity between mood and anxiety disorders, our sample is a strength. Anxiety disorders will be more difficult to identify when youth also have significant mood pathology, enabling us to evaluate the SCARED under realistic and challenging clinical conditions (Zhou et al., 2002).

All youth in the sample were evaluated with a rigorous diagnostic assessment that included a semi-structured interview with both the youth and parent followed by a consensus diagnosis by a licensed clinical psychologist or child psychiatrist, providing the opportunity to test diagnostic efficiency of the SCARED-P against DSM-IV-TR diagnoses using receiver operating characteristic (ROC) analyses. Additionally, we explored whether age or sex moderated SCARED-P performance. We collected SCARED parent report in order to focus on a younger age-group than could complete self-report efficiently, because some anxiety disorders are more common in childhood, and because parent report is more practical for large-scale screening, taking into account the need to collect parent consent before youth report could be collected in school, primary care, or community settings (Mineka, Watson, & Clark, 1998). Previous studies have found mixed results for the level of agreement between parent and child report on the SCARED (Cosi et al., 2010; Dirks et al., 2014; Wren et al., 2004).

Based on earlier studies (Birmaher et al., 1999; Cosi et al., 2010; Gardner et al., 2007), we expected the SCARED-P total and subscale scores to discriminate youth with anxiety disorder diagnoses from youth with other diagnoses. Additionally, we expected the SCARED-P total score to outperform the subscales for identifying *any* anxiety disorder, and the subscales to outperform the total score for predicting matched specific diagnoses (e.g., generalized anxiety disorder [GAD] subscale with GAD diagnosis). We anticipated that the AUCs would be *different* for girls and boys based on previous studies showing that girls tend to score higher than boys on the SCARED-P (Boyd, Ginsburg, Lambert, Cooley, & Campbell, 2003; Essau, Muris, & Ederer, 2002), but we did not hypothesize about the direction of the difference. We also expected that child age would affect performance of the SCARED-P, with better performance found when parents rate anxiety in older youth (Hale et al., 2011; Wren et al., 2004). Finally, we estimated multilevel diagnostic likelihood ratios (DLRs; Jaeschke et al., 1994) for ranges of scores on the SCARED total and subscale scores, and provided a clinical vignette to illustrate how to apply DLRs to aid assessment for an individual case.

Method

Participants and procedures

The current study used baseline data from the Longitudinal Assessment of Manic Symptoms (LAMS) study, which was designed to investigate the long-term trajectory of youth presenting for treatment with elevated symptoms of mania (ESM), regardless of whether they met criteria for a bipolar spectrum disorder. ESM has been conceptualized as a transdiagnostic dimension of emotion dysregulation, and one of the key goals for LAMS has been to investigate how this emotional dysregulation plays out over the course of development. Participants were recruited from nine outpatient mental health clinics

associated with the four LAMS university partners. Two cohorts of children were enrolled: youth aged 6 to 12.9 with ESM (as indicated by a score 12 or greater on the Parent General Behavior Inventory 10-item mania scale (PGBI-10M; Youngstrom, 2008), along with a demographically matched subset of youth aged 6 to 12.9 with scores less than 12 on the PGBI-10M. Due to primary aims of the LAMS study, all youth with scores greater than 12 on the PGBI-10M were invited to participate ($n=621$) while only a small percentage of youth with scores lower than 12 on the P-GBI ($n=86$) were invited. Horwitz et al. (2010) provide additional details about the recruitment strategy. All procedures were approved by the IRB at each participating university. Written informed consent from parents/guardians and assent from participating children were obtained. A total of 707 children enrolled between December 2005 and December 2008. The demographic characteristics of the sample are reported in Table 1.

Trained research assistants interviewed all youth and their parent/caregiver using the Scheduled for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (KSADS-PL-W; Findling, 2010; Geller, Williams, Zimmerman, & Frazier, 1996; Kaufman et al., 1997) to assess for both current and past psychiatric symptoms. All diagnoses followed unmodified DSM-IV-TR diagnostic criteria. Initial rater training occurred during a three-day start-up meeting, followed by rating along with taped interviews, and then by leading administrations of the assessment instruments with a reliable rater scoring along. To prevent rater drift following training, interviewers rated taped administrations of the K-SADS-PL-W and mood severity ratings. The κ for K-SADS-PL-W psychiatric diagnoses was 0.82 (Findling et al., 2010). A licensed clinical psychologist or child psychiatrist reviewed all diagnoses and KSADS data. SCARED-P scores were not considered when determining the diagnoses.

Measures

KSADS-PL-W—KSADS-PL-W (Findling, 2010; Geller et al., 1996; Kaufman et al., 1997) was administered to children and their parent or caregiver to assess for current and past psychiatric symptoms. The KSADS-PL-W is a widely used semi-structured interview that demonstrates strong reliability; in the present study inter-rater reliability for K-SADS-PL-W diagnoses was 0.82 (Findling et al., 2010).

PGBI-10M—PGBI-10M (Youngstrom, 2008) was administered to parents as part of the baseline assessment. The PGBI-10M is an empirically-derived scale carved from the Parent General Behavior Inventory, keeping items that best discriminate between bipolar disorder and other diagnoses. The PGBI-10M is highly internally consistent in the present sample, $\alpha=.92$. Though the PGBI-10M was designed to identify youth at risk for bipolar disorder, it does not include all symptoms associated with bipolar disorder, and those symptoms it does assess may also be associated with other childhood disorders.

SCARED-P—SCARED-P (Birmaher et al., 1999) was completed by caregivers as part of the baseline assessment. The SCARED-P consists of 41 items that assess child anxiety symptoms for the last 3 months. Each item is rated on a 3-point Likert type scale: 0 (*not true or hardly ever true*) to 2 (*true or often true*). Total scores range from 0 to 82, with high scores

indicating more anxiety symptoms; total scores of 25 or above are often used as the clinical cut-off (Birmaher et al., 1999; Monga et al., 2000). The SCARED-P has a five-factor structure corresponding with subscales for GAD, panic disorder, separation anxiety, social phobia, and school avoidance. The present analyses excluded the school avoidance subscale because it does not correspond to DSM diagnostic criteria, and because some studies have found that a four factor model, without school avoidance, fits better (Cosi et al., 2010). The current study examined diagnostic efficiency of total SCARED-P scores as well as scores for the GAD, and separation anxiety scales. In the present sample Cronbach's $\alpha = .95$ for the Total score, $\alpha = .90$ for the generalized anxiety subscale, and $\alpha = .85$ for the separation anxiety subscale.

Analytic Plan

Nonparametric estimates of the area under the curve (AUC) from receiver operating characteristic (ROC) analyses quantified diagnostic efficiency of the SCARED-P total and subscale scores. A rough guideline for evaluating AUC values is: $< .70 =$ poor, $.70 - .79 =$ fair; $.80 - .89 =$ good; and $.90 - 1.00 =$ excellent (Swets, 1988), although values higher than .90 in mental health contexts are often the result of design flaws such as comparing clinical cases to healthy controls (Youngstrom et al., 2006). AUCs were calculated for the target condition of *any* anxiety disorder using both the SCARED-P total and SCARED-P subscale scores, in order to evaluate whether total or subscales scores were better able to discriminate youth with any anxiety disorder (including GAD, panic disorder, OCD, PTSD, anxiety disorder NOS, separation anxiety, specific phobia, and social phobia) from other youth in the sample. Additionally, AUCs were calculated for GAD and separation anxiety using both SCARED-P total and SCARED-P subscales scores (matched with the respective diagnosis). Panic disorder was diagnosed in only 1% of the sample and social phobia was diagnosed in only 3%; because there were so few youth diagnosed with panic disorder or social phobia these diagnostic categories could not be evaluated independently, though these youth were included in the broad anxiety disorder category (Kraemer, 1992). Venkatraman's permutation test compared ROC curves (Venkatraman, 2000; Venkatraman & Begg, 1996). When a binary diagnostic test is being compared with a continuous test, this test has more statistical power than alternatives to test differences across AUCs, as it examines equality of the two ROC curves at all operating points, rather than just comparing the two overall AUC estimates (Venkatraman, 2000). Specifically, this method evaluates equality of the entire ROC curves by examining the absolute difference of the two ROC curves with a reference permutation distribution of this difference (Venkatraman & Begg, 1996).

Consequently, a significant result indicates existence of asymmetry at an operating classification point regardless of AUCs. In other words, this allows for identification of differences in diagnostic accuracy between two predictors, even when the point-estimate overall AUCs are the same. Moderator analyses tested whether diagnostic efficiency for the SCARED-P total and subscale scores changed significantly depending on whether or not the youth had a mood disorder diagnosis. Additionally, moderator analyses examined differences due to demographics, comparing males and females, and youth under 11 to those 11 and over, as previous studies have found age and sex differences on parent ratings on the SCARED (Birmaher et al., 1997; Wren, Bridge & Birmaher, 2004). Though the SCARED-P

has been validated among diverse samples (Gonzalez et al., 2012; Hale et al., 2011), we also tested whether results differed between White and non-White youth.

Finally, we calculated diagnostic likelihood ratios (DLRs) for cut-points yielding the best combination of sensitivity and specificity from ROC curves (Robin et al., 2011). DLRs based on optimal cut-points provide clinically useful information for predicting likelihood of a diagnosis. DLRs of less than 1.0 indicate that the observed score is associated with lower than average odds; DLRs of 1.0 mean that the score does not change the odds; DLRs between 2 and 5 are a small increase of the odds and potentially clinically meaningful; DLRs between 5.0 and 10.0 are a moderate increase, and DLRs greater than 10 are often clinically decisive (Straus et al., 2005).

The sampling design for LAMS under-sampled low scorers on the PGBI-10M by design. We weighted the observed cases to account for both probability of selection (all cases scoring 12+, along with a small matched sample of low scorers), as well as any differences in response rate, given that a family's participation was voluntary. Sample weights accounted both for the unequal probability of selection (by design) and any potential demographic differences introduced by non-response. First, the inverse of the probability of selection became a base weight. Second, logistic regression estimated the probability of agreeing to participate in the LAMS study once invited. The final model included the main effect of site, ESM status, age, sex, race, Hispanic ethnicity and insurance status, and all two-way interactions with ESM (except for Hispanic ethnicity due to small cell size). The distribution of predicted probabilities created 10 propensity cells. The non-response weight was the inverse of the unweighted response rate per propensity cell (Little & Vartivarian, 2003). The final weight was the product of the base weight and the non-response weight, which was rescaled such that the sum was equal to the sample size.

All analyses used SPSS-Version 22.0 or pROC package in R. Missing data were minimal (range 0-7% at the scale or summary variable level) and the missing at random assumption was tenable, so each analysis used listwise deletion.

Results

Demographics

Table 1 presents demographic characteristics of participants with and without anxiety disorders. Among the 707 participants, 31% of children were diagnosed with one or more anxiety disorders. Prevalence rates for GAD and separation anxiety were both 9%, $n=22$ youth met criteria for both separation anxiety and GAD. There were no age or sex differences between youth with and those without an anxiety disorder. The group with no anxiety disorders included a higher percentage of African American children. The SCARED-P total and SCARED-P subscale scores were higher in children with anxiety disorders, as expected. There were no differences on the PGBI-10M between those with versus without an anxiety disorder.

As expected, the SCARED-P total score was highly correlated with the scale scores ($r_s=.82-.85$, $p_s<.001$). The SCARED-P total and scale scores were moderately correlated with baseline P-GBI-10M scores ($r_s=.34-.42$, $p_s<.001$).

Diagnostic Efficiency Statistics

We examined back-to-back histograms for all distributions to visually inspect for degenerate distributions or outliers that might be problematic for ROC analysis (Youngstrom, 2014). AUCs for the total score were fair to good (.69–.89; see Table 2). Venkatraman's test compared AUCs for the SCARED-P total versus SCARED-P subscale scores across target anxiety conditions. Subscale scores outperformed the total score for predicting GAD (AUC = .86, $p<.0005$), but not for separation anxiety (AUC = .89, $p=.343$). The differences, though significant on the Venkatraman permutational test, were small enough that it was important to examine differences in the DLRs to decide if there might be clinical utility in interpreting the subtests.

Moderator analyses indicated that presence of a mood disorder changed performance of the SCARED-P. Among youth who *did not* have a comorbid mood disorder, compared to those with mood disorder, the GAD subscale performed better at predicting GAD ($p=.013$).

Moderator analyses identified significant interactions with sex; the GAD subscale performed better in males than females ($p=.016$); see Table 3. When we compared scale performance between children younger than 11 and those 11 and older, there were no differences; see Table 4.

When testing moderation across race/ethnicity, the SCARED-P total score performed better for non-White youth for predicting any anxiety disorder ($p=.013$), and GAD ($p=.005$). Additionally, the GAD subscale score did better predicting GAD among non-White youth ($p=.041$). There was no difference predicting separation anxiety by either total score or separation anxiety subscale.

Diagnostic likelihood ratios were computed using cut-off scores designed to find the optimal balance between sensitivity and specificity (Robin et al., 2011); see Table 5. There was some degeneracy in the data, due to the fact that some youth with high scores on the SCARED-P did not have anxiety disorders, often due to a mood disorder (population pyramid graphical display available from authors upon request). Consequently, we found that two DLRs – one indicating low risk, the other indicating elevated risk – best represented the relation between scale score and diagnosis in this sample. More fine-grained segments resulted in non-linear changes in DLR estimates with increasing SCARED-P scores due to high scores in cases with mood disorders absent an anxiety disorder. Using optimal cut-off scores, a SCARED-P total score of 22 or above was associated with a moderate increase in the odds of receiving one or more anxiety disorder diagnoses (DLR=2.71). In contrast, total scores of 21 and less were associated with reduced likelihood of meeting diagnostic criteria for any anxiety disorder (DLR=0.55). The GAD and separation anxiety subscale scores produced at least a moderate increase in the odds of predicting matched anxiety disorders; GAD subscale 8+, DLR=3.55, separation anxiety subscale 8+, DLR=5.35. Scores below the optimal cut-off for the two subscale scores were associated with reduced likelihood of an anxiety diagnosis, see

Table 5. Scores below the optimal cutoff are useful to rule-out specific diagnoses of anxiety disorders while scores above the optimal cut-off may help to identify the most likely anxiety diagnosis. All optimal cutoff scores yielded acceptable levels of sensitivity (.82) and specificity (.79), except the sensitivity (.56) for the total score predicting *any* anxiety disorders.

Discussion

Results indicated that individuals' scores on the SCARED-P are statistically valid for discriminating cases with anxiety disorders from a comparison group of treatment-seeking youth diagnosed with a wide range of disorders, including mood disorders, ADHD, and disruptive behavior disorders. Findings replicated prior work and indicated that elevated scores on the SCARED-P scales are associated with anxiety diagnoses. The AUC of .69 is similar to the .67 estimate reported by Monga et al. (2000) for the total score identifying cases with any anxiety disorder in another clinical sample.

We extended prior work by (a) evaluating the SCARED-P in the largest outpatient sample investigated to date, (b) testing performance of both the total score and GAD and separation subscale scores for identifying any anxiety disorder, as well as (c) GAD and separation anxiety, (d) directly comparing the performance of scales head-to-head in the same sample, and (e) testing whether sex, age, comorbid mood disorder, or race statistically moderated the diagnostic efficiency of scales. As hypothesized, we found that the total score performed well for the global purpose of identifying the presence of any anxiety disorder, whereas the GAD subscale showed a small but significant advantage for identifying GAD. We also found evidence that sex moderated the performance of some SCARED scales, although the difference in performance we found was subtle enough that though statistically significant, it was not clinically compelling.

This is the first paper to report diagnostic likelihood ratios for the SCARED. These are person-level effect sizes, whereas AUC is a group level effect size. An analogy would be that DLRs are to AUC as Jacobson's clinically significant change would be to Cohen's d : both DLR and clinically significant change apply effect to individual cases instead of group summaries. DLRs that represent *low* or *elevated* risk for an anxiety disorder are not intended to be diagnostic, rather, they provide a guideline for clinicians to consider in the context of other information relevant to the diagnostic decision making process. It may seem counterintuitive that a score in the elevated range would result in a very different DLR from a score just a few points away in the low risk range, but in the context of a screening tool, such information is likely to help clinicians. It helps determine whether further assessment for anxiety disorders is necessary – the clinician still must decide whether an anxiety disorder is present (Youngstrom, 2012).

Incorporating actuarial methods as part of the assessment process enables clinicians to integrate multiple sources of data, improve specificity of predictions made about diagnosis and prognosis, and limit the influence of bias and heuristics that tend to cloud clinical judgment (Aegisdottir et al., 2006; Dawes, Faust, & Meehl, 1989; Grove et al., 2000). To

illustrate this process using a nomogram – a simple, paper-and-pencil tool for combining likelihood ratios – we provide a clinical vignette below.

Clinical Illustration

Roger is a 10-year-old African American boy whose mother, Cecilia, called for an appointment because of “a lot of irritability and difficulty sleeping.” Before the appointment, the clinic sent a packet of information, forms, and symptom scales to be filled out and brought to the appointment, including a SCARED-P. When Roger and Cecilia arrived for their appointment, the clinic receptionist scored the scales; when the receptionist gave you the scored scales, she noted that the most elevated scores were on the SCARED-P. Cecilia’s answers yielded scores of 26 (total), 10 (GAD), 9 (separation anxiety), 6 (social phobia), and 1 (panic). You also see that on a questionnaire about family history of psychiatric illness, Cecilia has endorsed anxiety. Using the probability nomogram (see Figure 2), you begin with a base rate for anxiety disorders of 20%, which is consistent with the referral pattern at your clinic (Youngstrom et al., 2014). Next, you incorporate the knowledge that Roger’s mother has an anxiety disorder, which increases Roger’s risk by about 5 times. Drawing a line from the base rate through the number 5, corresponding to a 5 times greater risk, the posterior probability of Roger having an anxiety disorder is 56%. Next, you put the new probability on the left side, and incorporate the diagnostic likelihood ratio of 2.71 associated with Roger’s total score of 26 on the SCARED-P. Drawing the line from 56%, through 2.71, we get a posterior probability of 78%. If we had more information *from distinct sources*, such as a self-report from Roger, we could continue to refine the posterior probability. The results of the nomogram exercise are not diagnostic, but they will guide next steps using a structured or semi-structured diagnostic interview.

When you meet him, Roger appears sullen and responds with monosyllabic answers. His mother reports that Roger worries a lot; she said that he often expresses fears that he will fail in school, though his grades are good, and he likes to know where she is at all times. Recently, he has been missing a lot of school, and, on the days he does attend, he spends most of the day in the nurse’s office. He complains of headaches and upset stomach, and asks the nurse repeatedly to call his mother to pick him up. Cecilia also mentions that Roger is a “bad sleeper” and often wakes her up during the night, wanting to know if she is OK. The two of them have not been getting along well; she describes him as being “rude” and as having a “hot temper,” especially when she tries to get him to go to school. Cecilia reports that she is very worried about him. Following your interview with Cecilia and Roger, you conclude that the chief complaint of irritability and difficulty sleeping appear to be due to separation anxiety with some GAD symptoms, confirming the probability incorporating the clinic base rate of anxiety disorders, Roger’s SCARED-P total score, and his family history of anxiety using the nomogram.

The key features of the process are that (a) the interpretation is based on empirical weights for the different factors (the DLRs), (b) it is rapid – connecting the dots on a nomogram takes a few seconds, and it would be similarly fast to enter the numbers in an online or smartphone application, (c) the interpretation is more consistent and accurate than clinicians would achieve using typical, impressionistic interpretation of the same information (Jenkins

et al., 2011; Straus, Glasziou, Richardson, & Haynes, 2011). Because the SCARED-P is available at no charge, there is no additional expense added to the assessment process. In contrast, a meta-analysis found the kappa for agreement between structured diagnostic interviews and unstructured clinical interviews about the presence of anxiety disorders was in the .26 to .34 range based on 3050 cases (Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009).

Though the SCARED-P does not discriminate cases with anxiety disorders from those with other disorders well enough to be used alone to make a diagnosis, it does have utility as a component of the assessment process. The present study adds to the data indicating that, in addition to identifying youth with anxiety in community samples, the SCARED-P can also help to identify youth with anxiety in clinical samples. This is important, as the ability to distinguish healthy youth from youth with anxiety is not as helpful as being able to distinguish youth with anxiety from youth with depression or other internalizing symptoms. In a typical clinical scenario, where a caregiver brings a child for assessment and treatment, the SCARED-P can help the clinician to determine whether an anxiety disorder is likely, or if the symptoms might be due to depression or another disorder. This can be particularly helpful for internalizing disorders, which may be more difficult to assess in young people (Kendall et al., 2010).

Additionally, results of this study suggest that the SCARED-P can help to distinguish between specific anxiety disorders, not just between anxiety disorders and other disorders more broadly, further helping to focus the diagnostic decision. In fact, the DLRs for elevated subscale scores are higher than the DLR for the total score (predicting *any* anxiety disorder). Related, subscale sensitivity scores are better, indicating that they are likely more accurate estimates of risk than the total score. Furthermore, scores below the suggested cutoffs are informative, suggesting that the likelihood of an anxiety disorder is *lower*, because scores that low are rare in cases with anxiety as compared to other diagnoses (as indicated by a DLR <1). Interestingly, though the subscales performed better overall, moderator analyses suggest that the SCARED-P total score may be less susceptible than the GAD subscale to inflated scores due to comorbid mood disorders. Because both anxiety (and in particular GAD) and mood disorders are associated with negative affect, the presence of a mood disorder can lead to a false positive score on screening tools for anxiety (e.g., Austin et al., 2004). Tools, such as the SCARED-P total score, that distinguish anxiety from mood disorders can be very useful, particularly when many presenting symptoms are germane to both mood and anxiety.

As shown above, diagnostic likelihood ratios can be integrated easily with other information to aid clinical diagnosis. Though cut-off scores are helpful, they do not provide adequate information to know *how much* an individual's diagnostic picture changes given a higher or lower score. Additionally, literature on the SCARED-P offers a range of suggestions for cutoff scores, which may limit clinical utility.

To our knowledge, this is the first study to examine clinical utility of the SCARED-P and to provide DLRs, based on clinical data, to aid clinicians who use the SCARED-P in their practice. Other strengths of the study include investigation of SCARED-P subscales, along

with the total score, to determine diagnostic efficiency for matched diagnoses (e.g., the SCARED-P GAD subscale matched with GAD diagnosis). Though psychometric data (factor analysis results, AUCs) are helpful when evaluating whether or not to use a measure, they provide less clinical utility. We hope that, by providing both evidence of SCARED-P's diagnostic efficiency and a guide for incorporating it into an evidence-based approach to assessment, that clinicians will gain comfort with this process.

Limitations

Our findings used consensus diagnoses as the criterion. It is likely that the performance of the SCARED-P would look significantly better if we based diagnoses only on the KSADS interview with the parent. This would narrow the sources of input into the diagnosis (Campbell & Fiske, 1959), eliminating trait variance measured by youth report, interviewer behavioral observations, or clinical judgment on the part of the professional leading the consensus process. We agree with those that argue that the incorporation of these other sources of information enhances the validity of the resulting diagnosis (Bossuyt et al., 2003; Garb, 1998; Spitzer, 1983), even though the apparent effect sizes may be smaller (e.g., Youngstrom, Genzlinger, Egerton, & Van Meter, in press, found that effect sizes were +.25 d larger for parent scales when the diagnostic criterion interview only included the parent). For anxiety disorders, in particular, the use of consensus diagnoses likely increases the generalizability of results, as few clinicians would treat anxiety in the youth without first interviewing the youth and integrating their perceptions with the caregiver report (Silverman & Ollendick, 2005).

Though the SCARED-P has clear clinical utility, previous studies suggest that youth self-report on the SCARED may provide better information about the likelihood of an anxiety diagnosis (Wren et al., 2004; Kraemer, 2003) among older samples. Some evidence indicates that children are better reporters of internalizing symptoms than their parents (Muris, Merckelbach, Van Brakel, Mayer, & Birgit, 1999). Unfortunately, only parent report was collected in the present study due to youth age. Future studies should compare both measures to see if one outperforms the other and whether there is incremental value for including both.

Our findings suggest that the SCARED-P might perform differently depending on the child's age and sex. To our knowledge, this is the first study to examine diagnostic efficiency of the SCARED-P among different age and sex groups, so it is difficult to say whether or not different cutoff scores and DLRs should be used depending on age and sex. Because we found one significant interaction, it will be important to examine sex differences in future studies of the SCARED to better understand its performance in different populations.

Importantly, our population was demographically diverse. Results of our study, and others, suggest that the SCARED-P can be used effectively across racial and ethnic groups (Canals et al., 2012; Cosi et al., 2010; DeSousa et al., 2013; Gonzalez et al., 2012; Hale et al., 2011; Hariz, 2013). In fact, in our sample, the SCARED-P tended to perform as well or better among non-White youth. Due to recruitment strategies, our sample may be less diagnostically diverse than some outpatient clinical samples. The rate of mood disorders was high; though this may affect the generalizability of our results, given the high rates of

comorbid anxiety and mood disorders, we believe that testing the diagnostic efficiency of the SCARED-P in this sample offers conservative estimates. In outpatient samples with lower rates of mood disorders, the SCARED-P is likely to perform as well, or better, given that mood disorders can create false positives on anxiety measures.

Though 221 youth met criteria for an anxiety disorder, too few youth met criteria for panic disorder ($n=5$) or social phobia ($n=20$) to evaluate the diagnostic accuracy of the SCARED-P subscales for these specific diagnoses. The youth with panic disorder or social phobia were included in the “any anxiety” category though, so we know that a score above 22 could be indicative of either of these disorders. Future studies, perhaps with slightly older samples and higher prevalence of these anxiety disorders, should evaluate performance of the SCARED-P panic and social phobia subscales.

Finally, the present study’s recruiting strategy made it necessary to weight the data in order to provide more generalizable results, as is done in epidemiological studies or surveys. The recruitment plan was designed strategically to meet primary goals of the original study and thus, fewer children negative for ESM were recruited. As with any study, we hope other investigators will seek to replicate results and build confidence in the generalizability of these findings.

We hope that other investigators will also publish diagnostic efficiency data for other measures. Although the SCARED-P has many positive qualities – it is free, easy-to-use, and has been validated in community, primary care, and clinical populations (Birmaher et al., 1997; Boyd et al., 2003; DeSousa et al., 2013; Gonzalez et al., 2012; Hale et al., 2005; Hariz, 2013), there are other measures that can also be used to assess for youth anxiety disorders (e.g., the Revised Children’s Anxiety and Depression Scale, the Multidimensional Anxiety Scale for Children). Being able to directly compare the effect sizes and/or how sample characteristics moderate scale performance can help clinicians to choose the best measure for their clinical setting. For example, the published AUC for the GAD subscale of the RCADS in an outpatient sample is .76 (Cohen’s $d=.96$; Chorpita, Moffitt, & Gray, 2005), suggesting that if GAD were the primary diagnosis of interest, the SCARED-P might be a better tool to use (AUC=.86, Cohen’s $d=1.53$; the effect size is significantly higher, $p<.01$, based on Hanley & McNeil’s test).

Conclusion

Current results suggest that elevated scores on the SCARED-P (total or subscales) are associated with a 3 to 5-fold increased likelihood that an individual meets criteria for an anxiety disorder. From a clinician’s perspective, this information can be very helpful in determining whether further assessment and/or treatment is warranted (Youngstrom et al., 2014) as well as when considering differential diagnoses, particularly when integrated with other sources of information (family history, base rates).

Acknowledgments

FUNDING

National Institute of Mental Health:
5R01MH073801-10,5R01MH073816-10,5R01MH073953-10,5R01MH073967-10

References

- Aegisdottir S, White MJ, Spengler PM, Maugherman AS, Anderson LA, Cook RS, Rush JD. 2006; The Meta-Analysis of Clinical Judgment Project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*. 34:341–382. DOI: 10.1177/0011000005285875
- Aschenbrand SG, Angelosante AG, Kendall PC. 2005; Discriminant validity and clinical utility of the CBCL with anxiety-disordered youth. *Journal of Clinical Child and Adolescent Psychology*. 34:735–746. DOI: 10.1207/s15374424jccp3404_15 [PubMed: 16232070]
- Austin AA, Chorpita BF, Chorpita BF. 2004; Temperament, anxiety, and depression: comparisons across five ethnic groups of children. *Journal of Clinical Child & Adolescent Psychology*. 33(2): 216–226. [PubMed: 15136185]
- Baxter AJ, Vos T, Scott KM, Ferrari AJ, Whiteford HA. 2014; The global burden of anxiety disorders in 2010. *Psychological Medicine*. 44:2363–2374. DOI: 10.1017/S0033291713003243 [PubMed: 24451993]
- Beesdo-Baum K, Knappe S, Fehm L, Höfler M, Lieb R, Hofmann SG, Wittchen HU. 2012; The natural course of social anxiety disorder among adolescents and young adults. *Acta Psychiatrica Scandinavica*. 126(6):411–425. DOI: 10.1111/j.1600-0447.2012.01886.x [PubMed: 22632172]
- Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, Baugher M. 1999; Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A Replication Study. *Journal of the American Academy of Child & Adolescent Psychiatry*. 38(10):1230–1236. DOI: 10.1097/00004583-199910000-00011 [PubMed: 10517055]
- Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, Neer SM. 1997; The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale Construction and Psychometric Characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*. 36(4):545–553. DOI: 10.1097/00004583-199704000-00018 [PubMed: 9100430]
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, de Vet HCW. 2003; Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *British Medical Journal*. 326(7379):41–44. DOI: 10.1136/bmj.326.7379.41 [PubMed: 12511463]
- Boyd RC, Ginsburg GS, Lambert SF, Cooley MR, Campbell KDM. 2003; Screen for Child Anxiety Related Emotional Disorders (SCARED): Psychometric Properties in an African-American Parochial High School Sample. *Journal of the American Academy of Child & Adolescent Psychiatry*. 42(10):1188–1196. DOI: 10.1097/00004583-200310000-00009 [PubMed: 14560168]
- Campbell DT, Fiske DW. 1959; Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*. 56:81–105. [PubMed: 13634291]
- Canals J, Hernández-Martínez C, Cosi S, Domènech E. 2012; Examination of a cutoff score for the Screen for Child Anxiety Related Emotional Disorders (SCARED) in a non-clinical Spanish population. *Journal of Anxiety Disorders*. 26:785–791. DOI: 10.1016/j.janxdis.2012.07.008 [PubMed: 23023158]
- Chorpita BF, Moffitt CE, Gray J. 2005; Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy*. 43(3):309–322. DOI: 10.1016/j.brat.2004.02.004 [PubMed: 15680928]
- Christon LM, McLeod BD, Jensen-Doss A. 2014; Evidence-based assessment meets evidence-based treatment: An approach to science-informed case conceptualization. *Cognitive and Behavioral Practice*. doi: 10.1016/j.cbpra.2013.12.004
- Clark LA, Watson D. 1991; Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*. 100(3):316–336. [PubMed: 1918611]
- Cosi S, Canals J, Hernández-Martínez C, Vigil-Colet A. 2010; Parent–child agreement in SCARED and its relationship to anxiety symptoms. *Journal of Anxiety Disorders*. 24(1):129–133. DOI: 10.1016/j.janxdis.2009.09.008 [PubMed: 19864109]
- Dawes R, Faust D, Meehl P. 1989; Clinical versus actuarial judgment. *Science*. 243(4899):1668–1674. DOI: 10.1126/science.2648573 [PubMed: 2648573]

- DeSousa D, Salum G, Isolan L, Manfro G. 2013; Sensitivity and specificity of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A community-based study. *Child Psychiatry & Human Development*. 44:391–399. DOI: 10.1007/s10578-012-0333-y [PubMed: 22961135]
- Dirks MA, Weersing VR, Warnick E, Gonzalez A, Alton M, Dauser C, Woolston J. 2014; Parent and youth report of youth anxiety: Evidence for measurement invariance. *Journal of Child Psychology and Psychiatry*. 55:284–291. DOI: 10.1111/jcpp.12159 [PubMed: 24552483]
- Essau CA, Muris P, Ederer EM. 2002; Reliability and validity of the Spence Children’s Anxiety Scale and the Screen for Child Anxiety Related Emotional Disorders in German children. *Journal of Behavior Therapy and Experimental Psychiatry*. 33(1):1–18. DOI: 10.1016/S0005-7916(02)00005-8 [PubMed: 12389796]
- Findling RL, Youngstrom EA, Fristad MA, Birmaher B, Kowatch RA, Arnold LE, McCue Horwitz S. 2010; Characteristics of children with elevated symptoms of mania: the Longitudinal Assessment of Manic Symptoms (LAMS) study. *Journal of Clinical Psychiatry*. 71:1664–1672. DOI: 10.4088/JCP.09m05859yel [PubMed: 21034685]
- Garb, HN. *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association; 1998.
- Gardner W, Lucas A, Kolko DJ, Campo JV. 2007; Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *Journal of the American Academy of Child & Adolescent Psychiatry*. 46:611–618. DOI: 10.1097/chi.0b013e318032384b [PubMed: 17450052]
- Geller, B, Williams, M, Zimmerman, B, Frazier, J. WASH-U-KSADS (Washington University in St Louis Kiddie Schedule for Affective Disorders and Schizophrenia). St Louis, MO: Washington University; 1996.
- Gonzalez A, Weersing VR, Warnick E, Scahill L, Woolston J. 2012; Cross-ethnic measurement equivalence of the SCARED in an outpatient sample of African American and Non-Hispanic White youths and parents. *Journal of Clinical Child & Adolescent Psychology*. 41(3):361–369. DOI: 10.1080/15374416.2012.654462 [PubMed: 22397682]
- Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, Kaplan K. 2004; Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry*. 61(8):807–816. DOI: 10.1001/archpsyc.61.8.807 [PubMed: 15289279]
- Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. 2000; Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*. 12(1):19–30. DOI: 10.1037/1040-3590.12.1.19 [PubMed: 10752360]
- Hale WW, Raaijmakers Q, Muris P, Meeus WIM. 2005; Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in the general adolescent population. *Journal of the American Academy of Child & Adolescent Psychiatry*. 44(3):283–290. DOI: 10.1097/00004583-200503000-00013 [PubMed: 15725973]
- Hale WW, Crocetti E, Raaijmakers QAW, Meeus WHJ. 2011; A meta-analysis of the cross-cultural psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED). *Journal of Child Psychology and Psychiatry*. 52(1):80–90. DOI: 10.1111/j.1469-7610.2010.02285.x [PubMed: 20662993]
- Hamilton JD. 2001; Do we under utilize actuarial judgement and decision analysis? *Evidence Based Mental Health*. 4(4):102–103. DOI: 10.1136/ebmh.4.4.102 [PubMed: 11721514]
- Hanley JA, McNeil BJ. 1983; A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 148:839–843. [PubMed: 6878708]
- Hariz N. 2013; Reliability and validity of the Arabic Screen for Child Anxiety Related Emotional Disorders (SCARED) in a clinical sample. *Psychiatry Research*. 209(2):222. doi: 10.1016/j.psychres.2012.12.002 [PubMed: 23312477]
- Hettema JM, Neale MC, Kendler KS. 2001; A review and meta-analysis of the genetic epidemiology of anxiety disorders. *American Journal of Psychiatry*. 158:1568–1578. [PubMed: 11578982]
- Horwitz SM. 2010; Longitudinal Assessment of Manic Symptoms (LAMS) study: background, design, and initial screening results. *Journal of Clinical Psychiatry*. 71:1511–1517. DOI: 10.4088/JCP.09m05835yel [PubMed: 21034684]

- Hunsley J, Elliott K, Therrien Z. 2014; The efficacy and effectiveness of psychological treatments for mood, anxiety, and related disorders. *Canadian Psychology/Psychologie canadienne*. 55(3):161–176. DOI: 10.1037/a0036933
- Jaeschke R, Guyatt GH, Sackett DL, Group, t. E.-B. M. W. Guyatt G, Bass E, Wilson M. 1994; Users' Guides to the Medical Literature. *JAMA: The Journal of the American Medical Association*. 271:703–707. DOI: 10.1001/jama.1994.03510330081039 [PubMed: 8309035]
- Jenkins MM, Youngstrom EA, Washburn JJ, Youngstrom JK. 2011; Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology: Research and Practice*. 42(2):121–129. DOI: 10.1037/a0022506
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Ryan N. 1997; Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*. 36(7):980–988. DOI: 10.1097/00004583-199707000-00021 [PubMed: 9204677]
- Kendall PC, Compton SN, Walkup JT, Birmaher B, Albano AM, Sherrill J, Piacentini J. 2010; Clinical characteristics of anxiety disordered youth. *Journal of Anxiety Disorders*. 24(3):360–365. DOI: 10.1016/j.janxdis.2010.01.009 [PubMed: 20206470]
- Kessler R, Berglund P, Demler O, Jin R, Merikangas K, Walters E. 2005; Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 62:593–602. DOI: 10.1001/archpsyc.62.6.593 [PubMed: 15939837]
- Kowatch R, Youngstrom E, Danielyan A, Findling R. 2005; Review and meta-analysis of the phenomenology and clinical characteristics of mania in children and adolescents. *Bipolar Disorders*. 7(6):483–496. DOI: 10.1111/j.1399-5618.2005.00261.x [PubMed: 16403174]
- Kraemer, HC. *Evaluating medical tests*. Thousand Oaks, CA: Sage Publications; 1992.
- Little RJ, Vartivarian S. 2003; On weighting the rates in non-response weights. *Statistics in Medicine*. 22(9):1589–1599. DOI: 10.1002/sim.1513 [PubMed: 12704617]
- Lynch FL, Clarke GN. 2006; Estimating the economic burden of depression in children and adolescents. *American Journal of Preventive Medicine*. 31(6, Supplement 1):143–151. DOI: 10.1016/j.amepre.2006.07.001
- Merikangas KR, He JP, Brody D, Fisher PW, Bourdon K, Koretz DS. 2010; Prevalence and Treatment of Mental Disorders Among US Children in the 2001–2004 NHANES. *Pediatrics*. 125(1):75–81. DOI: 10.1542/peds.2008-2598 [PubMed: 20008426]
- Mineka S, Watson D, Clark LA. 1998; Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*. 49:377–412. DOI: 10.1146/annurev.psych.49.1.377
- Monga S, Birmaher B, Chiappetta L, Brent D, Kaufman J, Bridge J, Cully M. 2000; Screen for child anxiety-related emotional disorders (SCARED): Convergent and divergent validity. *Depression and Anxiety*. 12(2):85–91. DOI: 10.1002/1520-6394(2000)12:2<85::AID-DA4>3.0.CO;2-2
- Muris P, Merckelbach H, Van Brakel A, Mayer, Birgit. 1999; The revised version of the screen for child anxiety related emotional disorders (scared-r): Further evidence for its reliability and validity. *Anxiety, Stress, & Coping*. 12(4):411–425. DOI: 10.1080/10615809908249319
- Nelson LJ, Rubin KH, Fox NA. 2005; Social withdrawal, observed peer acceptance, and the development of self-perceptions in children ages 4 to 7 years. *Early Childhood Research Quarterly*. 20(2):185–200. DOI: 10.1016/j.ecresq.2005.04.007
- Norman GR, Eva KW. 2010; Diagnostic error and clinical reasoning. *Medical Education*. 44(1):94–100. DOI: 10.1111/j.1365-2923.2009.03507.x [PubMed: 20078760]
- Phillips BM, Lonigan CJ, Driscoll K, Hooe ES. 2002; Positive and negative affectivity in children: A multitrait-multimethod investigation. *Journal of Clinical Child & Adolescent Psychology*. 31(4):465–479. [PubMed: 12402566]
- Rapee RM. 2012; Family Factors in the Development and Management of Anxiety Disorders. *Clinical Child and Family Psychology Review*. 15:69–80. DOI: 10.1007/s10567-011-0106 [PubMed: 22116624]
- Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. 2009; Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews.

- International Journal of Methods in Psychiatric Research. 18(3):169–184. DOI: 10.1002/mpr.289 [PubMed: 19701924]
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. 2011; pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 12:77.doi: 10.1186/1471-2105-12-77 [PubMed: 21414208]
- Sakolsky D, Birmaher B. 2008; Pediatric anxiety disorders: management in primary care. *Current Opinion in Pediatrics*. 20(5):538–543. DOI: 10.1097/MOP.0b013e32830fe3fa [PubMed: 18781116]
- Seligman LD. 2011; Cognitive-behavioral therapy for anxiety disorders in youth. *Child and Adolescent Psychiatric Clinics of North America*. 20(2):217–238. DOI: 10.1016/j.chc.2011.01.003 [PubMed: 21440852]
- Shankman S, Lewinsohn P, Klein D, Small J, Seeley J, Altman S. 2009; Subthreshold conditions as precursors for full syndrome disorders: A 15-year longitudinal study of multiple diagnostic classes. *Journal of Child Psychology and Psychiatry*. 50(12):1485–1494. DOI: 10.1111/j.1469-7610.2009.02117.x [PubMed: 19573034]
- Silverman WK, Ollendick TH. 2005; Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*. 34:380–411. DOI: 10.1207/s15374424jccp3403_2 [PubMed: 16026211]
- Spitzer RL. 1983; Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry*. 24(5):399–411. [PubMed: 6354575]
- Straus, SE, Glasziou, P, Richardson, WS, Haynes, RB. Evidence-based medicine: How to practice and teach EBM. 4th. New York, NY: Churchill Livingstone; 2011.
- Swets J. 1988; Measuring the accuracy of diagnostic systems. *Science*. 240:1285–1293. DOI: 10.1126/science.3287615 [PubMed: 3287615]
- Tomb M, Hunter L. 2004; Prevention of anxiety in children and adolescents in a school setting: the role of school-based practitioners. *Children & Schools*. 26(2):87–101. DOI: 10.1093/cs/26.2.87
- Van Meter A, Youngstrom E, Youngstrom JK, Ollendick T, Demeter C, Findling RL. 2014; Clinical decision making about child and adolescent anxiety disorders using the Achenbach System of Empirically Based Assessment. *Journal of Clinical Child and Adolescent Psychology*. doi: 10.1080/15374416.2014.883930
- Venkatraman ES. 2000; A permutation test to compare receiver operating characteristic curves. *Biometrics*. 56:1134–1138. [PubMed: 11129471]
- Venkatraman ES, Begg CB. 1996; A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. *Biometrika*. 83(4):835–848. DOI: 10.1093/biomet/83.4.835
- Wren FJ, Bridge JA, Birmaher B. 2004; Screening for childhood anxiety symptoms in primary care: Integrating child and parent reports. *Journal of the American Academy of Child & Adolescent Psychiatry*. 43:1364–1371. DOI: 10.1097/01.chi.0000138350.60487.d3 [PubMed: 15502595]
- Youngstrom EA. 2008; Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. *Journal of Clinical Psychiatry*. 69:831–839. [PubMed: 18452343]
- Youngstrom EA. 2012; Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child & Adolescent Psychology*. 42(1):139–159. DOI: 10.1080/15374416.2012.736358 [PubMed: 23153181]
- Youngstrom EA. 2014; A primer on Receiver Operating Characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*. 39:204–221. DOI: 10.1093/jpepsy/jst062 [PubMed: 23965298]
- Youngstrom EA, Choukas-Bradley S, Calhoun CD, Jensen-Doss A. 2014; Clinical guide to the Evidence-Based Assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice*. 22:20–35. DOI: 10.1016/j.cbpra.2013.12.005
- Youngstrom EA, Genzlinger J, Egerton G, Van Meter AC. Multivariate meta-analysis of the discriminative validity of caregiver, youth, and teacher rating scales for pediatric bipolar disorder: Mother knows best about mania. *Archives of Scientific Psychology*.

- Youngstrom EA, Meyers OI, Youngstrom JK, Calabrese JR, Findling RL. 2006; Comparing the effects of sampling designs on the diagnostic accuracy of eight promising screening algorithms for pediatric bipolar disorder. *Biological Psychiatry*. 60:1013–1019. DOI: 10.1016/j.biopsych.2006.06.023 [PubMed: 17056395]
- Zhou, XH, Obuchowski, NA, McClish, DK. *Statistical methods in diagnostic medicine*. New York, NY: Wiley; 2002.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

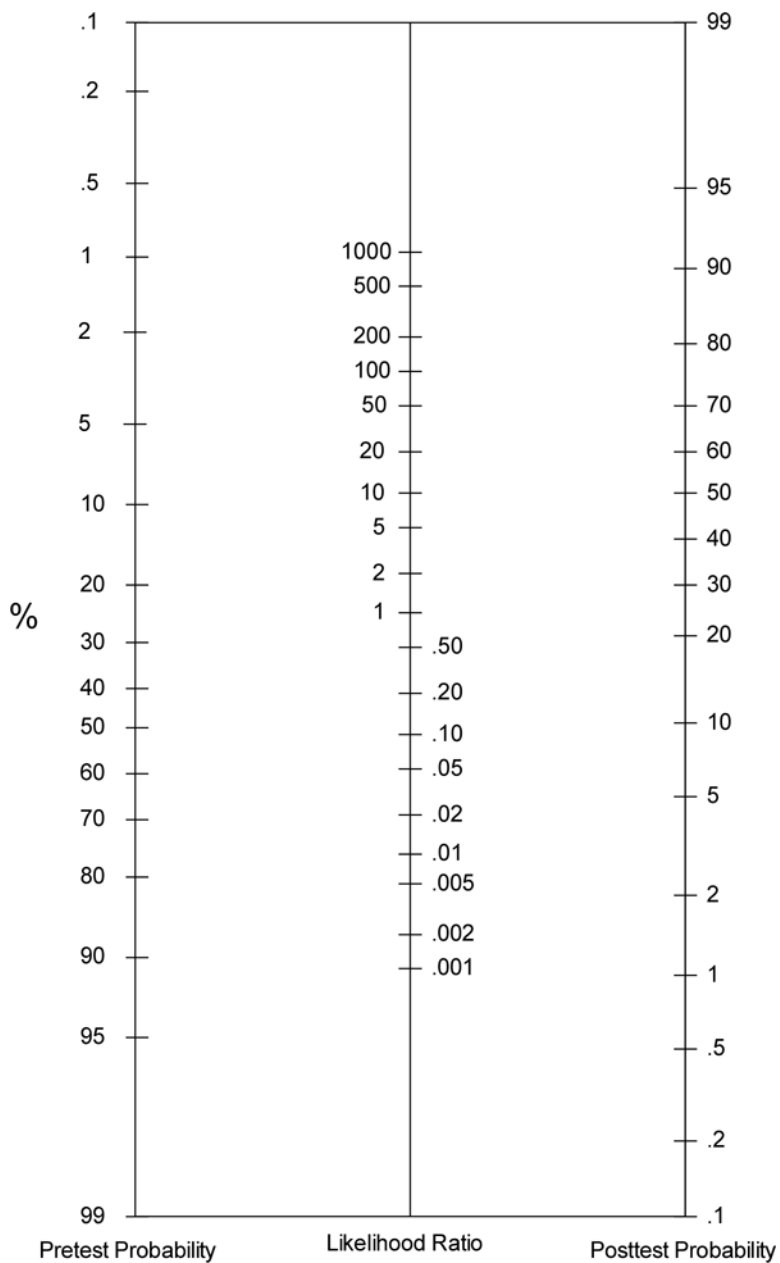


Figure 1. Nomogram for integrating likelihood ratios to determine a posterior probability
 Note: On the left side of the nomogram, the clinician would input the pretest probability – usually the base rate of the disorder under consideration – on the middle line, the likelihood ratio that corresponds with the first piece of clinical data (e.g., self report score, family history) is added. A line can then be drawn from the number on the pretest probability line through the likelihood ratio to the right side of the nomogram. The point at which the line intersects on the right side is the posttest probability – the new probability that the client has the disorder after integrating the base rate and one piece of clinical information. If there are additional pieces of information to add, the posttest probability becomes the new pretest probability, is moved to the left side, and the process begins again.

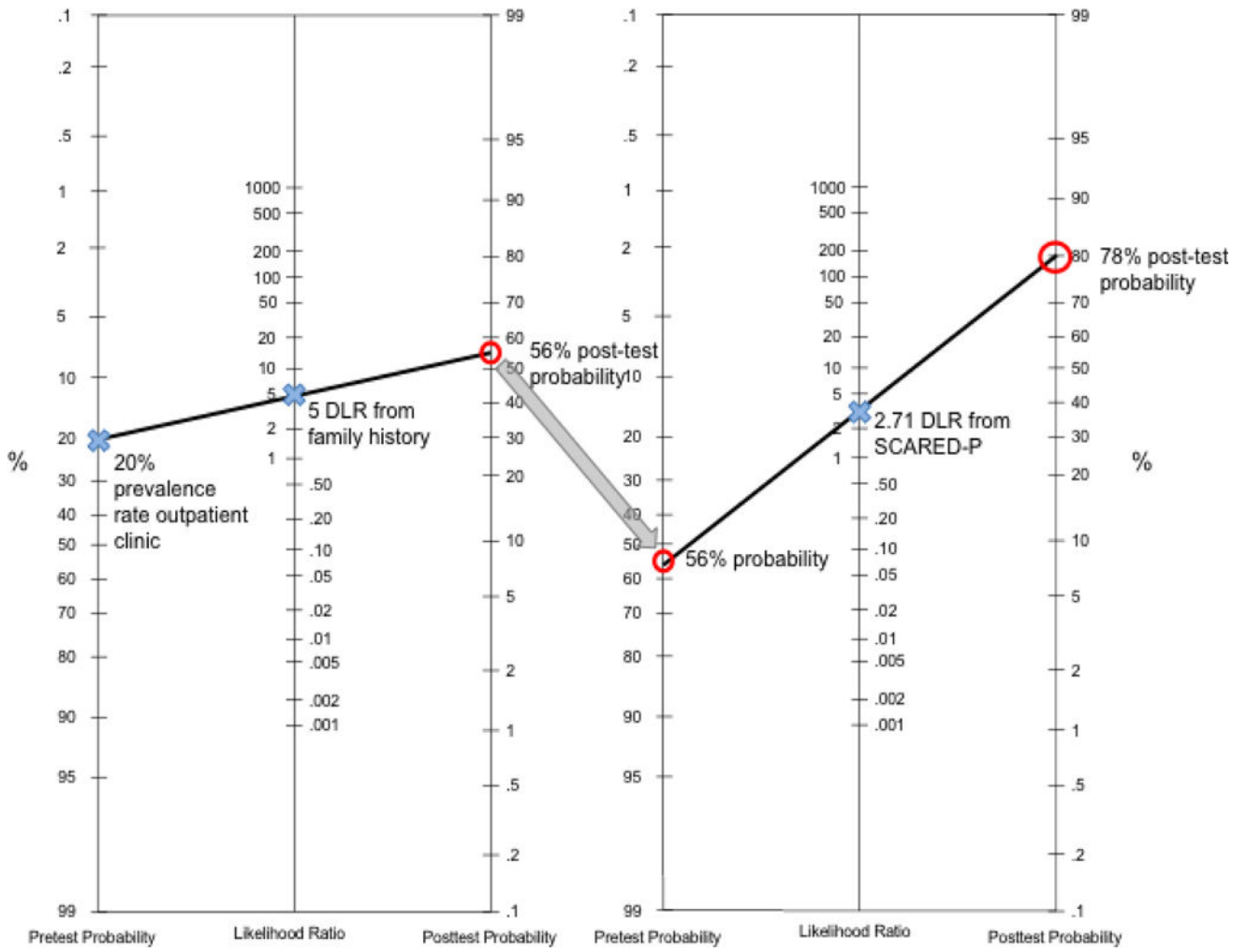


Figure 2.
Nomogram Example Using the SCARED-P for a youth seen at an outpatient clinic

Table 1

Demographic and Clinical Information

Variable	Any Anxiety (<i>n</i> = 221)	No Anxiety (<i>n</i> = 486)	Test Statistic	<i>p</i>
Age in years (<i>Mean, SD</i>)	9.3 (1.9)	9.5 (2.0)	<i>t</i> (705) = 1.10	.27
Male	<i>n</i> = 142 (64%)	<i>n</i> = 336 (69%)	$\chi^2(1) = 1.65$.20
Race ^a (White)	<i>n</i> = 163 (36%)	<i>n</i> = 292 (64%)	$\chi^2(4) = 21.10$	<.001
Depression	<i>n</i> =61 28%	<i>n</i> =63 13%	$\chi^2(1) = 22.51$	<.001
Bipolar spectrum disorder	<i>n</i> =49 22%	<i>n</i> =113 23%	$\chi^2(1) = .10$.75
Number of DSM-IV-TR Axis I diagnoses (<i>Mean, SD</i>)	3.5 (1.3)	2.0 (1.0)	<i>t</i> (338)=-15.60*	<.001
GAD	<i>n</i> = 66 (9%)	0		
Separation Anxiety	<i>n</i> = 65 (9%)	0		
Mean (<i>SD</i>)				
SCARED-P Total ¹	26.2(15.3)	14.7 (11.4)	<i>t</i> (305) = 9.55*	<.001
SCARED-P GAD ²	7.8 (4.8)	4.5 (3.8)	<i>t</i> (331) = 8.89*	<.001
SCARED-P Separation Anxiety ³	6.2 (4.4)	3.3 (3.4)	<i>t</i> (318) = 8.49*	<.001
P-GBI-10M ⁴	13.15 (7.05)	12.76 (7.23)	<i>t</i> (690) = -0.66	.51

* variances not equal, Levene's Test *p*<.001

^aComplete racial breakdown: 455 White youth, 187 Black youth, 61 biracial youth, 2 Asian, and 2 American Indian youth.

¹Screen for Child Anxiety Related Emotional Disorders-Parent rated total score;

²Screen for Child Anxiety Related Emotional Disorders-Parent rated generalized anxiety scale score;

³Screen for Child Anxiety Related Emotional Disorders-Parent rated separation anxiety scale score;

⁴Parent General Behavior Inventory 10-item mania scale

Table 2

ROC¹ Analyses Comparing SCARED-P Total Score² with SCARED-P Subscales³

Target Diagnosis	SCARED-P Score	95% Confidence Interval			Difference between AUCs ⁴
		Area Under Curve	Lower	Upper	
Any Anxiety	Total	.69***	.65	.73	
	GAD Subscale	.70***	.67	.74	<i>p</i> < .005
	Separation Anxiety Subscale	.65***	.61	.69	<i>p</i> < .005
GAD	Total	.80***	.76	.85	
	GAD Subscale	.86***	.83	.90	<i>p</i> < .005
Separation Anxiety	Total	.88***	.84	.92	
	Separation Anxiety Subscale	.89***	.86	.92	<i>p</i> = .371

p < .005, two tailed.

¹ Receiver Operating Characteristic;

² Screen for Child Anxiety Related Emotional Disorders-Parent rated total score;

³ Screen for Child Anxiety Related Emotional Disorders-Parent rated generalized anxiety, separation anxiety;

⁴ Area Under the Curve

Note: The Venkatraman method of assessing differences between ROC curves tests the difference at every operating point, which enables it to determine if one scale performs better than another, even if their overall AUCs are similar (Venkatraman, 2000). This can result in statistically significant differences between two scales, though the confidence intervals for the AUCs overlap.

Table 3
 ROC¹ Analyses for SCARED-P Total² and Subscale Scores³ Comparing Male and Female

Target Diagnosis and Predictor Subscale	Sex	95% Confidence Interval			Difference between AUCs ⁴
		Area Under Curve	Lower	Upper	
<i>SCARED-P Total score</i>					
Any Anxiety	Female <i>n</i> ⁵ =68	.66 [*]	.58	.75	<i>p</i> =.339
	Male <i>n</i> ⁵ =135	.70 ^{****}	.66	.74	
GAD	Female <i>n</i> ⁵ =25	.69 [*]	.57	.81	<i>p</i> =.048
	Male <i>n</i> ⁵ =40	.83 ^{****}	.80	.87	
Separation Anxiety	Female <i>n</i> ⁵ =19	.84 ^{****}	.77	.91	<i>p</i> =.302
	Male <i>n</i> ⁵ =42	.89 ^{****}	.84	.94	
<i>SCARED-P subscale scores</i>					
GAD	Female	.73 ^{**}	.64	.82	<i>p</i> =.024
	Male	.90 ^{****}	.87	.94	
Separation Anxiety	Female	.84 ^{****}	.75	.93	<i>p</i> =.447
	Male	.91 ^{****}	.87	.94	

* *p*<.05,

** *p*<.005,

*** *p*<.005, two tailed.

¹ Receiver Operating Characteristic;

² Screen for Child Anxiety Related Emotional Disorders-Parent rated total score;

³ Screen for Child Anxiety Related Emotional Disorders-Parent rated generalized anxiety, separation anxiety;

⁴ Area Under the Curve;

Weighted n
5

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

ROC¹ Analyses for SCARED-P Total² and Subscale Scores³ Comparing Youth under 11 to youth 11 Years and Older

Table 4

95% Confidence Interval					
Target Diagnosis	Age Group	Area Under Curve	Lower	Upper	Difference between AUCs ⁴
<i>SCARED-P Total score</i>					
Any Anxiety	<11 years <i>n</i> ⁵ =162	.74 ***	.70	.78	<i>p</i> =.899
	11+ years <i>n</i> ⁵ =41	.72 ***	.63	.81	
GAD	<11 years <i>n</i> ⁵ =55	.81 ***	.77	.86	<i>p</i> =.832
	11+ years <i>n</i> ⁵ =10	.85 **	.74	.97	
Separation Anxiety	<11 years <i>n</i> ⁵ =48	.87 ***	.81	.92	<i>p</i> =.606
	11+ years <i>n</i> ⁵ =13	.87 ***	.77	.97	
<i>SCARED-P subscale scores (paired with respective diagnosis)</i>					
GAD	<11 years	.88 ***	.84	.92	<i>p</i> =.754
	11+ years	.85 **	.75	.94	
Separation Anxiety	<11 years	.89 ***	.84	.93	<i>p</i> =.481
	11+ years	.96 ***	.93	.99	

** *p*<.005,

*** *p*<.005, two tailed.

¹ Receiver Operating Characteristic;

² Screen for Child Anxiety Related Emotional Disorders-Parent rated total score;

³ Screen for Child Anxiety Related Emotional Disorders-Parent rated generalized anxiety, panic disorder, separation anxiety, and social phobia scale scores;

⁴ Area Under the Curve;

⁵ Weighted n

Table 5
 Diagnostic Likelihood Ratios Based on Optimal Cut-off Scores for Detection of Anxiety Disorders

Target Diagnosis / Predictor Subscale	Score Range			
	Low Risk	Elevated Risk	Sensitivity	Specificity
Any Anxiety / Total Score	21	22+	.56	.81
Generalized Anxiety / GAD Subscale	DLR	0.55		
	Score	7	.82	.79
Separation Anxiety / Sep. Anx. Subscale	DLR	0.25		
	Score	7	.84	.85
	DLR	0.23		
		5.35		