

**PHS PUBLIC ACCESS**

Author manuscript

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:

J Clin Child Adolesc Psychol. 2017 ; 46(4): 511–522. doi:10.1080/15374416.2015.1030754.**Comparing the Diagnostic Accuracy of Five Instruments for Detecting Posttraumatic Stress Disorder in Youth****Dokyoung S. You,**

Department of Psychology, Texas A&M University

Eric A. Youngstrom,

Department of Psychology, University of North Carolina at Chapel Hill

Norah C. Feeny,

Department of Psychological Sciences, Case Western Reserve University

Jennifer Kogos Youngstrom, and

Department of Psychology, University of North Carolina at Chapel Hill

Robert L. Findling

Department of Psychiatry, Johns Hopkins University.

Abstract

OBJECTIVE—To compare diagnostic accuracy of five posttraumatic stress disorder (PTSD) measures in a large outpatient sample of youths aged 11 to 18 years.

METHOD—Index tests included a parent report (a rationally-derived scale from the Child Behavioral Checklist), a teacher report (the Teacher Report Form), and three youth reports—a PTSD scale from the Youth Self Report (YSR), Child PTSD Symptom Scale, and Child and Adolescent Trauma Survey. Interviews with the youth and caregiver using Schedule for Affective Disorders and Schizophrenia for School-age Children generated criterion diagnoses of PTSD. Diagnoses were blind to scores on the index tests.

RESULTS—Based on consensus diagnoses ($N=458$), 10% of youth had PTSD. Area under the curve (AUC) from receiver operating characteristic analyses and multilevel likelihood ratios evaluated test performance. All youth reports (AUCs .67-.73) outperformed the teacher report (AUCs .42-.48) at identifying PTSD. The YSR outperformed the caregiver reports (AUCs .57-.58). Combining tests did not improve prediction of PTSD. The YSR predicted PTSD even after controlling for a self-reported traumatic event, but checklist ratings of traumatic events had no incremental value after controlling for YSR scores. When a youth endorsed few symptoms, the likelihood of the youth having PTSD was low. Very high scores on the YSR were associated with a moderate increase in the likelihood of PTSD diagnosis.

Correspondence concerning this article should be addressed to Eric A. Youngstrom, Department of Psychology, University of North Carolina at Chapel Hill, CB #3270, Davie Hall, Chapel Hill, NC 27599-3270. eay@unc.edu.
(dokyoungyou@gmail.com)
(norah.feeny@case.edu)
(jkyoungstrom@gmail.com)
(RFindli1@jhmi.edu)

The other authors have no potential conflicts of interest to disclose.

CONCLUSION—The YSR appeared to be a useful diagnostic aid for youth PTSD, and could facilitate differential diagnosis of youth PTSD in outpatient settings.

Keywords

PTSD; diagnostic accuracy; Child Behavioral Checklist; evidence based assessment

Posttraumatic stress disorder (PTSD) affects 5% of the general adolescent population in the United States (Merikangas et al., 2010) and can occur in 3% to 49% of cases in pediatric mental health outpatient clinics (Gabbay, Oatis, Silva, & Hirsch, 2004; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). Despite its prevalence, it often goes undiagnosed. A recent meta-analysis found that only 3% of outpatient cases were clinically diagnosed with PTSD in samples where structured diagnostic interviews identified 9% of the cases as having PTSD (Rettew et al., 2009). Untreated childhood PTSD is linked to life-long psychological and physical health problems and to increased use of medical services (Goenjian et al., 2005; Mulvihill, 2005; Walker et al., 1999). Identifying PTSD is crucial to reduce its negative impact on health and social costs.

The American Academy of Child and Adolescent Psychiatry (AACAP) Practice Parameter recommends that clinicians routinely screen for trauma exposure and PTSD symptoms during an intake assessment, gathering information from both caregivers and youths to improve diagnostic accuracy (Cohen et al., 2010). Primary caregivers' perspective is essential when working with prepubertal youths (Scheeringa, Wright, Hunt, & Zeanah, 2006). Two key questions are what measures should be used as diagnostic aids, and whether the use of multiple informants improves assessment accuracy. Selecting instruments for pediatric PTSD is challenging, because few studies directly test discriminative validity against semi-structured or consensus diagnoses--considered the gold standard to diagnose PTSD (Cohen et al., 2010).

Also, to our knowledge, no study has examined teacher report as a way of identifying cases with PTSD in youths, although the items from several rationally derived PTSD scales could be extracted from teacher report versions of scales. For example, there are promising PTSD-oriented scales (Wolfe, 2007) that consist of items selected from the caregiver and youth forms of the Achenbach System of Empirically Based Assessment (ABSEBA; Achenbach & Rescorla, 2001). The teacher report version of the ASEBA includes the same items, but they have not been tested yet for identifying PTSD. On one hand, teacher report is often less sensitive to internalizing problems (De Los Reyes & Kazdin, 2005); but on the other hand, teacher report could be useful in situations where youth or parent report might be compromised (McDermott, 1994). Teacher report also could help identify cases where PTSD is associated with impairment across multiple settings.

When assessing PTSD in youth, clinicians should select instruments that are valid and clinically meaningful. According to Evidence Based Medicine (EBM), tests should be combined with other risk factors and clinical findings to revise probability estimates for diagnoses (Straus, Glasziou, Richardson, & Haynes, 2011). Evidence-Based Assessment (EBA) strategies (Youngstrom, 2013) have incorporated these ideas for anxiety disorders (Van Meter, Youngstrom, Youngstrom, Ollendick, & Findling, 2014), attention-deficit/

hyperactivity disorder (Frazier & Youngstrom, 2006), and pediatric bipolar disorder (Youngstrom et al., 2004; Youngstrom, Meyers, Youngstrom, Calabrese, & Findling, 2006), among other conditions.

This study's purpose was to compare diagnostic accuracy of five different instruments for PTSD in an outpatient sample aged 11 to 18 years. We hypothesized that the scales would show statistically significant discrimination between cases with versus without PTSD, and that the effect sizes would be medium to large by conventional standards, but moderate for clinical decision-making purposes based on effect sizes seen in prior reports for PTSD (e.g., Gillihan, Aderka, Conklin, Capaldi, & Foa, 2013) as well as using similar scales for anxiety disorders (Van Meter et al., 2014). Additional goals were to compare the diagnostic value of youths', caregivers', and teachers' reports and to examine whether combining multiple informants' reports would add significant diagnostic information. This study also examined whether endorsement of a traumatic event on a self-report checklist would provide additional diagnostic information. To our knowledge, this study is the first to compare diagnostic accuracy of proposed PTSD content scales derived from youth report or teacher report on the Achenbach System of Empirically Based Assessment (ASEBA) scales (Achenbach & Rescorla, 2001). Prior studies have examined criterion correlations or group differences, but not diagnostic accuracy or clinical interpretation of these scales (Foa, Johnson, Feeny, & Treadwell, 2001; Wolfe, 2007; Wolfe, Gentile, & Wolfe, 1989). We also explored whether teacher report provided useful information. Our last goal was to develop multi-level diagnostic likelihood ratios for those tests that demonstrate statistical significance to facilitate their use in clinical settings (Straus et al., 2011).

Method

Participants

Institutional Review Boards from Case Western Reserve University and Applewood Centers reviewed and approved all procedures. We recruited 458 participants from both places. Legal guardians provided written informed consent, and youth provided separate assent. Participants were a consecutive case series seeking mental health outpatient services located in an urban city for their youths from 2002 to 2008.

Youths were on average 13.5 years old ($SD = 1.9$; female 47%). Majority were African American (68%) followed by White, non-Hispanic (25%), Hispanic (2%), and other (5%). Caregivers were 77% biological parents, 9% grandparents, and 5% adoptive parents; 92% female, and a few (6%) had at least a college degree. Median income was between \$10,000 and \$ 14,999.

Inclusion criteria were youth aged between 11 and 18 years, with both caregiver and youth able to communicate in English. Exclusion criteria were pervasive developmental disorders or cognitive disability. The age of 11 was chosen as the cut-off because normative data for many self-report measures begin at that age, and youths of that age typically have sufficient cognitive development and reading ability to provide reliable responses (Achenbach & Rescorla, 2001). All participants completed the diagnostic interview.

Measures

Reference standard: Semi-structured Diagnostic Interview using the Schedule for Affective Disorders and Schizophrenia for School-Age Children (KSADS)—

Families completed a semi-structured interview with the KSADS-Present and Lifetime version (Kaufman et al., 1997). The KSADS includes a trauma-screening interview that formally established the presence of a clinically significant traumatic event and a PTSD module inquiring about symptoms associated with the event. PTSD diagnoses followed strict DSM-IV-TR criteria (American Psychiatric Association, 2000). Analyses used presence/absence of current diagnoses of PTSD, regardless of comorbidity.

Two-dozen interviewers were trained until inter-interviewer agreement exceeded $\kappa=.85$ at the item level and 1.0 at diagnostic level (see Youngstrom et al., 2005 for more details). One interviewer administered the KSADS sequentially with both the caregiver and the youth, and a different interviewer gathered the rating scales; diagnoses were blind to the rating scales. When the caregiver and youth provided discrepant information, the interviewer could re-interview either or both, and used best clinical judgment to reconcile the information. Reliability of KSADS diagnoses was $\kappa=.91$ compared to expert review of all available data from the interviews, family history, and prior treatment history.

Child Behavioral Checklist (CBCL)—The primary caregiver completed the 2001 version of the Achenbach CBCL (Achenbach & Rescorla, 2001). The CBCL is a widely used and standardized broad-coverage measure in clinical and research settings. The CBCL contains 118 items rated from 0 (*not true*) to 2 (*very true or often true*). Wolfe and colleagues rationally derived a PTSD scale based on item content, first proposing a 20-item scale ($\alpha=.89$) and then a 14-item scale ($\alpha=.86$) (Wolfe, 2007; Wolfe et al., 1989). We examined both the 20 (CBCL-20, $\alpha=.82$) and 14 item scale (CBCL-14, $\alpha=.79$ in our data). We also analyzed the CBCL Internalizing *T*-score as a benchmark for comparison. The Achenbach scales are all commercially distributed, with commercial scoring software. The 14-item scale is included in the output of the current version of the commercial software (Achenbach & Rescorla, 2007). Therefore, our analysis focused on the 14-item scale with the 20-item scale included for comparison.

Youth Self Report (YSR; Achenbach & Rescorla, 2001)—Adolescent participants completed the YSR, which contains the same PTSD-relevant items as the CBCL. The current study examined both the YSR-14 ($\alpha=.85$) and 20 item scales ($\alpha=.86$).

Teacher Report Form (TRF; Achenbach & Rescorla, 2001)—Teachers completed the TRF. The current study explored both the TRF-14 ($\alpha=.79$) and 20 item scales ($\alpha=.84$).

Child PTSD Symptom Scale (CPSS; Foa et al., 2001)—The CPSS is freely available and a widely used measure in clinical practice and research. This youth-report assesses the most distressing event and 17 DSM-IV PTSD symptoms over the past two weeks on a Likert scale, ranging from “*not at all* (0)” to “*5 or more times a week* (3).” The total score, ranging from 0 to 51, indicates PTSD symptom severity. The CPSS also includes seven yes/no functional impairment items. The impairment score ranges from 0 to 7 with higher scores

indicating more impairment. This study examined diagnostic accuracy of the CPSS with ($\alpha=.91$) and without the functional items ($\alpha=.92$). Prior studies found that the CPSS had high internal consistency ($\alpha=.89$) and discriminative validity (Foa et al., 2001; Gillihan et al., 2013). The CPSS is in the public domain, and easily scored by hand.

Child and Adolescent Trauma Survey (CATS; March, 1999)—The CATS combines a self-report of lifetime trauma history (yes/no) and current symptoms (0: *never* to 3: *often*) that has been used in several NIH funded projects. Different from traditional statistical modeling techniques, Item Response Theory analyses guided item selection ($\alpha=.86$; Suliman, Kaminer, Seedat, & Stein, 2005). We used the 12 PTSD symptom severity ratings ($\alpha=.90$). Suliman et al. (2005) examined the CATS' diagnostic accuracy in South African adolescents (AUC=.81). The present study is the first to test its discriminative validity in a USA sample, as well as using treatment-seeking cases to form the comparison group.

Procedure

After giving written consent and assent, youth participants and their families completed the KSADS interview. The clinical interview and self-reports were completed on the same day. When the youth was doing the interview, the parent/guardian completed the rating scales; the youth did the scales independently while the caregiver completed the KSADS. Youths and caregivers did not have access to each other's responses on the rating scales. The TRF was completed by one teacher that the family identified as having the most extensive contact with the youth. Diagnoses were blind to the rating scales. Participants received \$20 compensation.

Statistical Methods

Missing Data—Completion rates for measures were: CPSS 97%, CATS 95%, YSR 92%, and CBCL 87%. The TRF was available for 28% of cases ($n=128$); it was not gathered if families were interviewed during the summer. Little's MCAR Chi-Square test was not significant ($\chi^2=54.97, p=.20$), suggesting that data were well suited to the multiple-imputation procedure. We imputed missing values at the scale level and the imputation model included all the clinical syndrome scores, the demographic characteristics, and the diagnostic status of the cases. Multiple imputation with 10 imputed datasets maximized the usable sample size and avoided the possible introduction of bias due to missing data (Schafer & Graham, 2002). Despite the large amount of missing teacher data, we checked for patterns in missingness, evaluated potential bias via bootstrapping methods, and also compared results to ROC analysis without imputation for TRFs. Rubin's formula corrected the standard errors (Rubin, 1987).

Diagnostic Accuracy—The primary criterion measure grouped youths into two categories: those with versus without PTSD regardless of comorbid Axis I diagnoses. Nonparametric estimates of the area under the curve (AUC) from receiver operating characteristic (ROC) analyses (Youngstrom, 2014) quantified the overall diagnostic accuracy of each test using the pROC R package (Robin et al., 2011). Confidence intervals computed for ROC curves used bootstrap resampling methods (2000 resamples) to provide robust standard errors (Chernick, 2008). A rough guideline is AUC values $< .70$ =poor, $.70$.

.79=fair; .80-.89=good; and .90-1.00=excellent (Swets, 1988). It should be noted that these benchmarks may be aspirational for measures evaluated under clinically realistic and generalizable conditions; using a comparison group where all cases are seeking treatment will lower the diagnostic specificity (Youngstrom et al., 2006). The reliability of the diagnosis also constrains the magnitude of the AUC estimates (Kraemer, 1992). DeLong's test of paired ROC curves (DeLong, DeLong, & Clarke-Pearson, 1988) with Rubin's pooling methodology (Rubin, 1987), compared the diagnostic accuracy of the different index tests.

For measures that showed best discriminative validity, we calculated diagnostic likelihood ratios; DLRs = [True Positive/(True Positive+False Negative)]/[False Positive/(False Positive +True Negative)]. DLRs change the posterior odds associated with a particular test score: DLRs <1.0 lower the odds of a PTSD diagnosis; ~1.0 = no change; 2-5 =odds increase perhaps clinically meaningfully; 5-10 =large increase; and >10=often clinically decisive odds changes (Straus et al., 2011). DLRs can be combined with the prior probability of a diagnosis to generate posterior predictive values, using Bayes' Theorem and a calculator or a probability nomogram (Figure 1) (Jenkins, Youngstrom, Washburn, & Youngstrom, 2011; Straus et al., 2011).

Incremental Value of Combining Index Tests: Logistic regression analyses examined whether combinations of the index tests provided any incremental value (Hosmer & Lemeshow, 2013). Measures providing incremental validity can be combined within an EBA framework. We also examined whether endorsement of a trauma on the CPSS would improve diagnostic efficacy.

Results

Participants and Descriptive Statistics

Table 1 presents diagnostic characteristics. Among the youths who endorsed any of the screening interview items on the KSADS, parents and youths agreed about the youth meeting criteria for PTSD $\kappa = .45$, $p = .001$. The kappa was .70 when including all the cases where both agreed that there was no history of trauma exposure. This is greater agreement than often found between parent and youth about internalizing disorders (e.g., Grills & Ollendick, 2003; DiBartolo & Grills, 2006; Hawley & Weisz, 2003). There were 10 cases where the parent interview indicated the presence of PTSD, but the youth report alone would not have met criteria, and there were 6 cases where the youth report met criteria for PTSD, and the parent report did not. McNemar's test indicated that there was no significant difference in giving one or the other informant more weight, $p = .454$. There were 20 cases where both informants agreed that criteria for PTSD were met, and two with missing data from the youth.

Of 38 youth with PTSD, 70% were female. All youths with KSADS diagnoses of PTSD met criteria for exposure to a traumatic event based on the interview. However, only 74% of youths with PTSD reported a traumatic event on the CPSS and the rest left this section blank; and 61% of youths without PTSD reported a traumatic event on the CPSS. If clinicians used the CPSS items in isolation to identify exposure to a clinically meaningful traumatic event, these would translate to a sensitivity of 74% and a specificity of 39%.

Consequently, using a checklist to screen for a traumatic event is not effective for predicting a PTSD diagnosis; the sensitivity indicates it would miss more than a quarter of cases, and the corresponding specificity is still meager. Small correlations were found between endorsed trauma and youth self-reports ($r_s = .22-.36$, $p_s < .001$), whereas youth endorsed trauma was not significantly related to either parents' or teachers' reports ($p_s > .42$). Among 286 endorsed traumatic events on the CPSS, the most frequently endorsed was death of someone close (10%), but only 14% of youth endorsing someone's death were diagnosed with PTSD. In contrast, sexual assault was reported by 11 youths, half of whom met full criteria for PTSD. A similar pattern was observed with the CATS. Of 297 endorsed worst events on the CATS, the most frequently endorsed item was "me or someone I know well 'got sick and almost died or died'" (13%), but only 13% of youth endorsing this item were diagnosed with PTSD, whereas over a third of 20 youths endorsing the 'sexually abuse' item were diagnosed with PTSD. In general, there was no association between endorsing potentially traumatic events on the CPSS or CATS and having associated impairment based on the KSADS interview: Both produced kappa $< .03$, and chi-squared < 1.5 , n.s.

Table 2 presents descriptive statistics for the index tests and severity of PTSD symptoms separately for PTSD present versus absent, along with correlations. Correlations among youth measures were large. Youth reports correlated moderately with caregiver report, but not with teacher report--consistent with typical levels of caregiver-youth agreement and roughly consistent with teacher-youth agreement (Achenbach, McConaughy, & Howell, 1987).

Diagnostic Accuracy Statistics

All youth and caregiver reported scales achieved statistical significance discriminating cases with PTSD from the other youths in the sample; teacher report did not attain statistical significance in any analysis. Figure 2 shows the ROC curves. Based on the KSADS PTSD diagnosis, the YSR-14 and -20 scales performed in the "fair" range at identifying PTSD (see Table 2 for AUCs). DeLong's test compared the paired AUCs: Both the YSR-14 and -20 scales performed better than the YSR Internalizing T -score ($\chi^2 > 2.09$, $p_s < .037$), and they were not different from each other ($p = .52$). The CPSS and CATS performed fairly and poorly respectively, but they did not differ significantly from the YSR scales ($\chi^2_s < 0.94$, $p_s > .354$). Test performance of youth reports overlapped with the caregiver reports ($\chi^2_s < 1.75$, $p_s > .089$), but youth reports were superior to teacher report (all $\chi^2_s > 2.89$, $p_s < .026$). Although statistically significant, the AUC estimates of diagnostic accuracy for the CBCL 14- and 20-item scales fell in the "poor" range, albeit better than the TRF Internalizing T scores ($\chi^2_s > 3.23$, $p_s < .023$). Finally, all teachers' reports performed poorly regardless of the use of imputed (Table 2) or original dataset (AUCs ranging from .41 to .47, all n.s.). Bootstrapping detected minimal bias for all scales (< 0.00013). Hanley and McNeil's (1983) test of AUCs from independent samples indicated that the CATS obtained a significantly lower AUC in the present sample than in Suliman et al.'s (2005) school sample, AUC of .69 versus .81, $p < .05$. This pattern of results is consistent with the present sample consisting entirely of youths seeking outpatient services, whereas most participants in Suliman et al. were typically functioning students. Including healthy cases in the comparison group

routinely yields much larger effective sizes and optimistic estimates of diagnostic accuracy (Zhou, Obuchowski, & McClish, 2002).

Evaluation of Combinations of Index Tests

Logistic regressions found that none of the scales significantly improved model fit after controlling for the most powerful index tests, such as the YSR-14 or -20 (p values of Wald tests and χ^2 change $>.05$), with one exception: CATS scores contributed significant incremental information after controlling for the YSR-20 (Wald=4.52, $p=.034$). Combining them still failed to improve the overall classification accuracy. Additional logistic regressions indicated that--after controlling for each of the promising child measures (YSR-14 and -20; CPSS) - self-reported traumas did not predict PTSD diagnosis ($p>.08$). In contrast, youth measures predicted PTSD diagnoses even after controlling for self-reported traumas ($p<.002$). Among youths with PTSD, symptom severities on the promising measures were not different between the groups with and without self-reported trauma ($t<1.05$, $p>.21$). Taken together, self-reported severity scores detected a PTSD diagnosis more reliably than endorsement of trauma history. Therefore, higher symptom severity on youth self-reports should be followed by a semi-structured interview or structured involving both the youth and the caregiver to probe for the presence of PTSD.

Optimize Thresholds for Diagnostic Accuracy

Table 3 presents multilevel DLRs for those tests showing promising AUCs. Odds of a PTSD diagnosis were lower (DLRs <1.0) when the symptom severity on the YSR-14 and -20 were less than 60th percentile in the sample. Very high scores (top 10%) on the YSR-14 and -20, and high scores (top 25%) on the CPSS, moderately increased the likelihood of PTSD diagnosis. Screening applications often prefer high sensitivity cutoff scores in order to detect the most PTSD cases. The YSR-20 cutoff score of 9+ (90% sensitivity) yielded the highest post-test probability of a youth having PTSD compared to the DLRs+ of the other youth reports with 90% sensitivity cutoff scores (Table 3). However, using this traditional approach of 90% sensitivity would not yield high positive predictive value, as it only had a DLR+ of 1.37. The EBA approach with multilevel DLRs would provide more clinically useful diagnostic information. Table 3 presents the probabilities associated with test scores in scenarios with PTSD base rates of 10% (a conservative rate for outpatient mental health) and 40% (an upper bound rate of what might be encountered in an outpatient clinic; Gabbay et al., 2004; Rettew et al., 2009) to provide a sense of how the scales might work in clinical practice.

Discussion

Despite practice parameters recommending that clinicians evaluate potential traumatic events and PTSD (Cohen et al., 2010), evidence suggests that clinical diagnoses frequently miss more than half of PTSD diagnoses that would be identified via more structured approaches to assessment (Jensen-Doss, Youngstrom, Youngstrom, Feeny, & Findling, 2014; Rettew et al., 2009; Wolfe, 2007). The current study is the first to compare directly the diagnostic accuracy of five potential diagnostic scales for pediatric PTSD. As we hypothesized, the scales showed statistically significant discrimination between cases with

versus without PTSD except teacher reports. The effect sizes would normally be considered medium to large if expressed as Cohen's *d*, but when recast as AUC values, they reflected poor to fair classification accuracy. Youths were better informants about their own PTSD symptoms, although they did not always acknowledge traumatic events without a semi-structured interview that specifically inquired about them. Results indicated that the YSR-14 and 20 item scales provided the two highest AUC values, but all child measures outperformed teacher ratings. Youth scales performed in what is conventionally considered the "fair" range in detecting PTSD, except for poor performance by the CATS. Although both caregiver and teacher ratings performed less well, only teacher ratings performed statistically worse than youth ratings.

Combining scales did not improve prediction of PTSD diagnosis. Regardless of self-reported trauma, youth report predicted PTSD, suggesting that severity scores provided diagnostic information for PTSD even when youth did not report a traumatic event. This is consistent with the clinical observation that it is not the occurrence of an event, per se, but rather the psychological response to it that is significant with regard to PTSD. Furthermore, using a checklist as the sole source of information about traumatic events could miss a large portion of cases with PTSD – almost a quarter of cases with confirmed diagnoses of PTSD did not endorse any item on the trauma checklists. High scores on youth symptom scales should trigger careful interviewing about both potential trauma and PTSD. The follow-up interview should involve both the youth and the primary caregiver. Although youth and caregiver agreed about PTSD diagnoses as well or better than typical in other samples, each informant identified a subset of cases with PTSD that would have been missed without direct interview of both.

Our results are consistent with previous findings that youth are better reporters of their internalizing symptoms (De Los Reyes & Kazdin, 2005; Reich, 2000). The better performance of youth versus caregiver report on the same instrument is consistent with clinicians' perception that youth report often may be more valid for internalizing problems (Loeber, Green, & Lahey, 1990), yet also stands in contrast to the body of evidence showing that caregiver report outperforms youth report for symptoms of mania and pediatric bipolar disorder (Youngstrom, Jenkins, Jensen-Doss, & Youngstrom, 2012), ADHD (Pelham, Fabiano, & Massetti, 2005), including analyses using these same data (Youngstrom et al., 2005). This reduces concerns about the possibility that the stronger association of youth report was due to design artifacts, because the same design produced contrasting results for other diagnoses. Among the youth reports, the YSR was a better diagnostic aid than the others for the purpose of identifying potential PTSD in our sample. The YSR-14, YSR-20, CPSS, and CATS achieved statistically similar performance, with the YSR-14 and 20 performing marginally better in the present data. Currently, computer-generated output of the YSR-14 is available and very high YSR-14 scores (18+) may provide more clinically useful information than other scales for detecting youth PTSD. A score of 18+ on the YSR-14 is associated with 3.6 times greater odds of PTSD. We recommend the YSR-14 to assess for PTSD because the YRS-14 had a similar AUC to the YSR-20, and the YSR-14 is available for the clinician without extra work. Using the YSR-14 provides a pragmatic advantage as it is embedded in a widely used broad-coverage instrument. Alternatively, the

CPSS might be useful in settings with cost constraints that preclude using the Achenbach YSR.

Strengths of the present study include the use of a relatively large sample where recruitment was unlikely to introduce bias vis-à-vis PTSD diagnoses, and with a clinically representative comparison group that included all other youths in the case series regardless of diagnosis (Bossuyt et al., 2003; Zhou et al., 2002). This study tested the diagnostic accuracy of the measures in a natural clinical setting where status in terms of trauma or PTSD were not known ahead of time (Bossuyt et al., 2003). Our design examined effectiveness under real-world clinical conditions, (a) including a wide range of severity of PTSD presentations, (b) allowing comorbidity, and (c) using a diagnostically heterogeneous comparison group that included anxiety disorders, depression, and psychotic disorders. These conditions are likely to generate false positive results on a putative PTSD assessment, and test performance is likely to be worse compared to scenarios using a healthy control group or a diagnostically distilled design (Youngstrom et al., 2006; Zhou et al., 2002). Our AUC values were significantly lower than those previously reported by Suliman et al. for the CATS, also consistent with the prior study including a large number of healthy controls in the comparison group. Therefore, our assessment in a natural clinical setting makes our findings more likely to replicate in outpatient settings. Additionally, we had an adequate number of PTSD cases ($n = 38$) in our sample. Kraemer (1992) recommends having at least 10 cases at each marginal position when evaluating diagnostic accuracy. We also used multiple methods to evaluate diagnostic accuracy (i.e., global estimates and multilevel likelihood ratios) and to provide clinically useful information about change in likelihood of diagnosis depending on assessment findings. Furthermore, we followed the STARD reporting guideline to ensure that all relevant information was presented in the current paper (Bossuyt et al., 2003). Finally, bootstrapping methods produced robust estimates of confidence intervals by adjusting for the bias (Chernick, 2008).

Limitations

Although we compared multiple scales from up to three different informants, there are several other recommended measures for assessing PTSD symptoms in youths that were not evaluated here (see Fletcher, 2007, for detailed review). When clinicians are selecting scales for clinical use, they need to consider the full set of scales available, although meta-analyses and head-to-head comparisons will be the most informative ways of comparing scales. Our sample also consisted of youths age 11 and older, consistent with the age norms for the YSR. Prior work has shown that parent report measures can identify cases with PTSD as young as preschool age (Scheeringa et al., 2006), and there are promising evidence based treatments for that age group as well (Scheeringa et al., 2007). Our findings that youth report performed similarly to caregiver report are likely to be specific to the age-range of the sample, as younger children have lower reading ability and are less psychologically minded than adolescents and adults – factors that can reduce the reliability and validity of youth report on checklists (Anastasi & Urbina, 1997). Although the current study found chance performance for teacher's ratings in detecting youth PTSD, low response rate (28%) reduced power and precision for evaluating the teacher scores. However, missing data analyses indicated that the missing pattern was not systematic, and teacher report is often associated

with internalizing disorders weakly if at all (De Los Reyes & Kazdin, 2005). It seems unlikely that the most sensitive, accurate teachers would also be less likely to respond. Another caveat is that study participants were demographically diverse, with most being African American and low income (with median family income being below the poverty line). Future research should examine whether these measures perform similarly in other ethnic groups. Also, DSM-5 changed the diagnostic criteria for PTSD to use four symptoms cluster instead of three (American Psychiatric Association, 2013). The new cluster of symptoms--negative alterations in cognitions and mood associated with traumatic events—contains symptoms mostly from the avoidance symptoms cluster in DSM-IV. Future studies should evaluate the diagnostic accuracy of the measures for youth PTSD using the DSM-5 diagnostic criteria.

Clinical Implications

In contrast to older concepts of sensitivity and specificity, diagnostic likelihood ratios make it possible to combine information from test results with other data about cases to form revised probability estimates (Straus et al., 2011; Youngstrom, Choukas-Bradley, Calhoun, & Jensen-Doss, 2014). These methods also help glean useful information from tests that are not powerful enough to be used in isolation for high stakes decisions. Illustration with a clinical example makes this clearer. The base rate of a disorder or condition at a clinic is a good starting benchmark for evaluation. In the present sample, ~10% of youth had PTSD. Therefore, the initial odds would be 0.11 [= 0.10/0.90]. If a youth reported a score of 18+ on the YSR-14, the change in odds for PTSD is 3.64 (Table 3). Then, the new odds are 0.40 [= 0.11×3.64], which yields a posterior probability of 29% [= $0.40/(0.40+1)$]. The posttest probability can be estimated using online tools (e.g., <http://easycalculation.com/statistics/post-test-probability.php>) or using the nomogram (Figure 1). On the nomogram, first, draw a straight line from the pretest probability (10% on the far left) to the calculated DLR (3.64 on the middle). Next, extend the straight line to the posttest probability (about 30% on the far right). The combination of the two points sets the slope of the line. If the base rate were 40%--at the high end of what is reported in some outpatient clinics (Gabbay et al., 2004), very high YSR-14 scores would yield a positive predictive power of 71%, meaning that 71% of youths with a score of 18+ would have PTSD when drawn from a clinic with that base rate. The difference in posterior probabilities, from 29% to 71% in this example, illustrates the influence of the starting probability on the predictive values. The DLR and nomogram approach makes this connection visible and allows the clinician and patient to see how adding new information, such as additional tests or risk factors, modifies the risk estimates (Straus et al., 2011). When clinicians do not know the base rate at their clinics, they can use a reported base rate at similar settings or start with a quantification of their clinical judgment. Using these methods produces large improvements in the accuracy and consistency of clinical interpretations (Jenkins et al., 2011), and increases agreement about the next clinical action that is indicated, even when the initial estimate is not accurate (Jenkins, Youngstrom, Youngstrom, Feeny, & Findling, 2012).

Current findings suggest several clinical recommendations. First, youths provide more clinically useful information for PTSD diagnosis than teachers, at least in the age range of 11 to 18 years. Therefore, clinicians should prioritize youth reports in diagnosing youth

PTSD when multiple informants complete rating scales – recognizing that at younger ages caregiver report may be the better option (Scheeringa et al., 2006). The YSR-14 scale performed well in the present data at detecting youth PTSD among the contenders compared here, offering a small but significant improvement compared to the YSR Internalizing scores. The Internalizing scale is not specific to PTSD, cutting across anxiety disorders (Ferdinand, 2008; Van Meter et al., 2014) and mood disorders (Ferdinand, 2008; Youngstrom, 2014). When a youth endorses few symptoms on the YSR-14 scale, the likelihood of the youth having PTSD diagnosis is low; and high scores moderately increase the odds. Hence, the YSR-14 can provide helpful information about risk of youth PTSD during the outpatient evaluation, although it will not be a decisive piece of information by itself. The CPSS could be useful in situations where clinicians do not have the YSR, do not have the resources to purchase it, nor have a reason for a global evaluation at the time of assessment; otherwise the YSR-14 provides more breadth of information. Combining other informants' ratings or additional youth measures with the YSR-14 or CPSS did not improve overall detection of PTSD. Finally, the YSR-14 and CPSS provided incremental information even after controlling for a self-reported trauma, but not vice versa. We recommend that clinicians sensitively inquire about trauma when YSR-14 scores are high, even if the youth did not indicate trauma as a presenting problem or on a checklist. Use of empirically validated instruments such as the YSR-14 or the CPSS, though not sufficient to establish a diagnosis of PTSD on their own, can play a valuable role in improving the detection of PTSD in youths. Validated instruments can provide a helpful piece of evidence in formulating clinical decisions about diagnosis and treatment as they assist clinicians in calculating and revising the diagnostic probability within an EBA framework (Youngstrom, Choukas-Bradley, Calhoun, & Jensen-Doss, 2014). We recommend the diagnostic procedure to be as follows; (a) start with a broad measure, (b) look at PTSD scale, (c) ask about abuse and trauma history and risk factors, and (d) follow up with semi-structured or structured interview that involves both the youth and the primary caregiver. Clinicians already using the Achenbach Youth Self Report need not supplement with a more specialized trauma or PTSD scale. Results also emphasize the importance of doing a semi-structured interview around traumatic events, as screening checklists were not an adequate substitute in this sample. The combination of a broad checklist and a brief semi-structured interview suggests that small refinements to common practices could produce large improvements in the detection of trauma and PTSD in youths.

Acknowledgments

Dr. Eric Youngstrom has consulted with Lundbeck and Otsuka. Dr. Findling receives or has received research support, acted as a consultant and/or served on a speaker's bureau for Alcobra, American Academy of Child & Adolescent Psychiatry, American Physician Institute, American Psychiatric Press, AstraZeneca, Bracket, Bristol-Myers Squibb, CogCubed, Cognition Group, Coronado Biosciences, Dana Foundation, Elsevier, Forest, GlaxoSmithKline, Guilford Press, Johns Hopkins University Press, Johnson and Johnson, Jubilant Clinsys, KemPharm, Lilly, Lundbeck, Merck, NIH, Neurim, Novartis, Noven, Otsuka, Oxford University Press, Pfizer, Physicians Postgraduate Press, Purdue, Rhodes Pharmaceuticals, Roche, Sage, Shire, Sunovion, Supernus Pharmaceuticals, Transcept Pharmaceuticals, Validus, and WebMD. Dr. Feeny receives book royalties from Guilford Press and funding from NIH.

This research was supported in part by NIH R01 MH066647 (PI: E. Youngstrom).

References

- Achenbach TM, McConaughy SH, Howell CT. Child/Adolescent behavioral and emotional problems: Implication of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 101:213–232. doi: 10.1037/0033-2909.101.2.213. [PubMed: 3562706]
- Achenbach, TM., Rescorla, LA. Manual for the ASEBA School-Age Forms & Profiles. University of Vermont; Burlington, VT: 2001.
- Achenbach, TM., Rescorla, LA. Manual for the ASEBA School-Age Forms & Profiles. University of Vermont; Burlington, VT: 2007.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th--Text Revision ed.. Author; Washington, DC: 2000.
- American Psychiatric Association.. Diagnostic and statistical manual of mental disorders. 5th ed.. Author; Washington, DC: 2013.
- Anastasi, A., Urbina, S. *Psychological Testing*. 7th ed.. MacMillan; New York: 1997.
- 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(7379):41–44. doi: 10.1136/bmj.326.7379.41. [PubMed: 12511463]
- Chernick, M. *Bootstrap methods: A practitioner's guide*. 2nd ed.. Wiley; Hoboken, NJ: 2008.
- Cohen JA, Bukstein O, Walter H, Benson SR, Chrisman A, Farchione TR, Issues AWGOQ. Practice parameter for the assessment and treatment of children and adolescents with posttraumatic stress disorder. *Journal of American Academy of Child and Adolescent Psychiatry*. 2010; 49(4):414–430.
- De Los Reyes A, Kazdin AE. Informant Discrepancies in the Assessment of Childhood Psychopathology: A Critical Review, Theoretical Framework, and Recommendations for Further Study. *Psychological Bulletin*. 2005; 131(4):483–509. doi: 10.1037/0033-2909.131.4.483. [PubMed: 16060799]
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44(3):837–845. [PubMed: 3203132]
- DiBartolo PM, Grills AE. Who is best at predicting children's anxiety in response to a social evaluative task? A comparison of child, parent, and teacher reports. *Journal of Anxiety Disorders*. 2006; 20:630–645. doi: 10.1016/j.janxdis.2005.06.003. [PubMed: 16112539]
- Ferdinand RF. Validity of the CBCL/YSR DSM-IV scales Anxiety Problems and Affective Problems. *Journal of Anxiety Disorders*. 2008; 22(1):126–134. doi: 10.1016/j.janxdis.2007.01.008. [PubMed: 17321103]
- Fletcher, KE. Posttraumatic Stress Disorder.. In: Mash, EJ., Barkley, RA., editors. *Assessment of Childhood Disorders*. 4 ed.. Guilford Press; New York, NY: 2007. p. 398-446.
- Foa EB, Johnson KM, Feeny NC, Treadwell KR. The child PTSD Symptom Scale: a preliminary examination of its psychometric properties. *Journal of Clinical Child Psychology*. 2001; 30(3): 376–384. doi: 10.1207/S15374424JCCP3003_9. [PubMed: 11501254]
- Frazier TW, Youngstrom EA. Evidence-Based Assessment of Attention-Deficit/Hyperactivity Disorder: Using Multiple Sources of Information. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2006; 45(5):614–620. doi: 10.1097/01.chi.0000196597.09103.25. [PubMed: 16670656]
- Gabbay, V., Oatis, MD., Silva, RR., Hirsch, GS. Epidemiological aspects of PTSD in children and adolescents Posttraumatic stress disorders in children and adolescents: Handbook. W.W. Norton; New York: 2004. p. 1-17.
- Gillihan SJ, Aderka IM, Conklin PH, Capaldi S, Foa EB. The Child PTSD Symptom Scale: psychometric properties in female adolescent sexual assault survivors. *Psychological Assessment*. 2013; 25(1):23–31. doi: 10.1037/a0029553. [PubMed: 22867010]
- Goenjian AK, Walling D, Steinberg AM, Karayan I, Najarian LM, Pynoos R. A prospective study of posttraumatic stress and depressive reactions among treated and untreated adolescents 5 years after a catastrophic disaster. *American Journal of Psychiatry*. 2005; 162(12):2302–2308. doi: 10.1176/appi.ajp.162.12.2302. [PubMed: 16330594]

- Grills AE, Ollendick TH. Multiple informant agreement and the anxiety disorders interview schedule for parents and children. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2003; 42:30–40. doi: 10.1097/00004583-200301000-00008. [PubMed: 12500074]
- 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]
- Hawley KM, Weisz JR. Child, parent, and therapist (dis)agreement on target problems in outpatient therapy: the therapist's dilemma and its implications. *Journal of Consulting and Clinical Psychology*. 2003; 71:62–70. doi: 10.1037/0022-006X.71.1.62. [PubMed: 12602426]
- Hosmer, DW., Lemeshow, S. *Applied Logistic Regression*. 3rd ed.. Wiley; New York: 2013.
- Jenkins MM, Youngstrom EA, Washburn JJ, Youngstrom JK. Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology: Research and Practice*. 2011; 42(2):121–129. doi: 10.1037/a0022506.
- Jenkins MM, Youngstrom EA, Youngstrom JK, Feeny NC, Findling RL. Generalizability of evidence-based assessment recommendations for pediatric bipolar disorder. *Psychological Assessment*. 2012; 24(2):269–281. doi: 10.1037/a0025775. [PubMed: 22004538]
- Jensen-Doss A, Youngstrom EA, Youngstrom JK, Feeny NC, Findling RL. Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *Journal of Consulting and Clinical Psychology*. 2014 doi: 10.1037/a0036657.
- 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 & Adolescent Psychiatry*. 1997; 36(7):980–988. doi: 10.1097/00004583-199707000-00021. [PubMed: 9204677]
- Kraemer, HC. *Evaluating medical tests: Objective and quantitative guidelines*. Sage; Newbury Park, CA: 1992.
- 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(2):136–143.
- March, JS. Assessment of pediatric posttraumatic stress disorder.. In: Saigh, P., Bremner, J., editors. *Posttraumatic stress disorder: a comprehensive text*. Allyn & Bacon; Needham Heights, MA: 1999. p. 199-218.
- McDermott, PA. *National profiles in youth psychopathology: Manual of Adjustment Scales for Children and Adolescents*. Edumetric and Clinical Science; Philadelphia, PA: 1994.
- 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(10):980–989. doi: S0890-8567(10)00476-4 [pii]. [PubMed: 20855043]
- Mulvihill D. The health impact of childhood trauma: an interdisciplinary review, 1997-2003. *Issues Compr Pediatr Nurs*. 2005; 28(2):115–136. doi: 10.1080/01460860590950890. [PubMed: 16006380]
- Pelham WE Jr, Fabiano GA, Massetti GM. Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*. 2005; 34:449–476. doi: 10.1207/s15374424jccp3403_5. [PubMed: 16026214]
- Reich W. Diagnostic interview for children and adolescents (DICA). *Journal of the American Academy of Child & Adolescent Psychiatry*. 2000; 39:59–66. [PubMed: 10638068]
- Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research*. 2009; 18(3):169–184. doi: 10.1002/mpr.289. [PubMed: 19701924]
- Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*. 2005; 29:615–620. [PubMed: 16254746]
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011; 12:77. doi: 10.1186/1471-2105-12-77. [PubMed: 21414208]

- Rubin, DB. Multiple imputation for nonresponse in surveys. Wiley; Hoboken, NJ: 1987.
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7(2):147–177. [PubMed: 12090408]
- Scheeringa MS, Wright MJ, Hunt JP, Zeanah CH. Factors affecting the diagnosis and prediction of PTSD symptomatology in children and adolescents. *American Journal of Psychiatry*. 2006; 163(4): 644–651. doi: 10.1176/appi.ajp.163.4.644. [PubMed: 16585439]
- Scheeringa MS, Salloum A, Arnberger RA, Weems CF, Amaya-Jackson L, Cohen JA. Feasibility and effectiveness of cognitive-behavioral therapy for posttraumatic stress disorder in preschool children: two case reports. *Journal of Traumatic Stress*. 2007; 20:631–636. doi: 10.1002/jts.20232. [PubMed: 17721975]
- Straus, SE., Glasziou, P., Richardson, WS., Haynes, RB. Evidence-based medicine: How to practice and teach EBM. 4th ed.. Churchill Livingstone; New York, NY: 2011.
- Suliman S, Kaminer D, Seedat S, Stein D. Assessing post-traumatic stress disorder in South African adolescents: using the child and adolescent trauma survey (CATS) as a screening tool. *Ann Gen Psychiatry*. 2005; 4(1):2. doi: 10.1186/1744-859X-4-2. [PubMed: 15845137]
- Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988; 240(4857):1285–1293. [PubMed: 3287615]
- Van Meter A, Youngstrom E, Youngstrom JK, Ollendick T, Demeter C, Findling RL. Clinical decision making about child and adolescent anxiety disorders using the Achenbach System of Empirically Based Assessment. *Journal of Clinical Child and Adolescent Psychology*. 2014 doi: 10.1080/15374416.2014.883930.
- Walker EA, Unutzer J, Rutter C, Gelfand A, Saunders K, VonKorff M, Katon W. Costs of health care use by women HMO members with a history of childhood abuse and neglect. *Archives of General Psychiatry*. 1999; 56(7):609–613. [PubMed: 10401506]
- Wolfe, VV. Child sexual abuse.. In: Mash, EJ., Barkley, RA., editors. Assessment of childhood disorders. 4th ed.. Guilford Press; New York: 2007. p. 685-748.
- Wolfe VV, Gentile C, Wolfe DA. The impact of sexual abuse on children: A PTSD formulation. *Behavior Therapy*. 1989; 20(2):215–228. doi: [http://dx.doi.org/10.1016/S0005-7894\(89\)80070-X](http://dx.doi.org/10.1016/S0005-7894(89)80070-X).
- 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(1):139–159. doi: 10.1080/15374416.2012.736358. [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*. 2014; 39:204–221. doi: 10.1093/jpepsy/jst062. [PubMed: 23965298]
- Youngstrom EA, Choukas-Bradley S, Calhoun CD, Jensen-Doss A. Clinical guide to the Evidence-Based Assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice*. 2014 doi: 10.1016/j.cbpra.2013.12.005.
- 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. doi: 10.1097/01.chi.0000125091.35109.1e. [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 OI, Demeter C, Kogos Youngstrom J, Morello L, Piiparinen R, Calabrese JR. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders*. 2005; 7(Special Issue: Pediatric Bipolar Disorder):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. doi: 10.1016/j.biopsych.2006.06.023. [PubMed: 17056395]

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

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

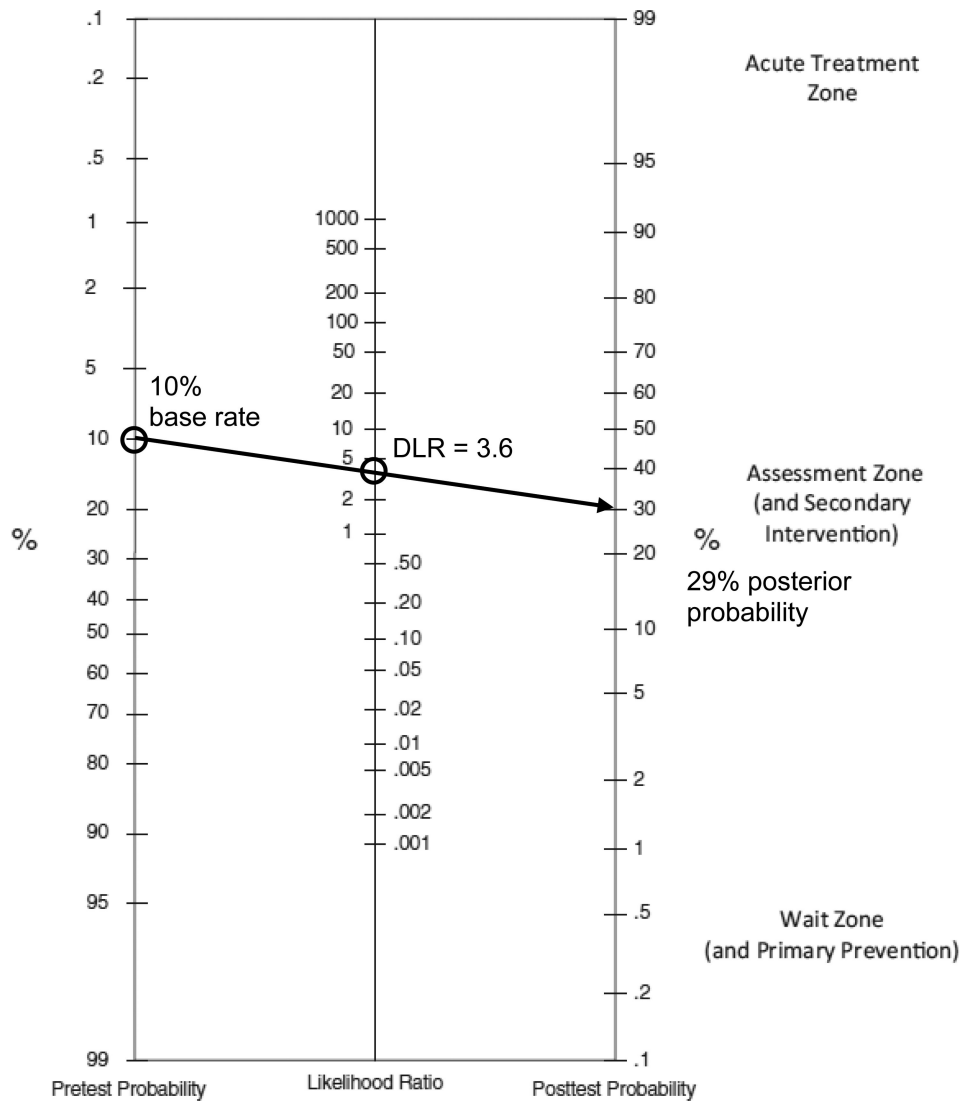


Figure 1. Probability nomogram for combining probability with diagnostic likelihood ratios presented by Straus et al. (2011)

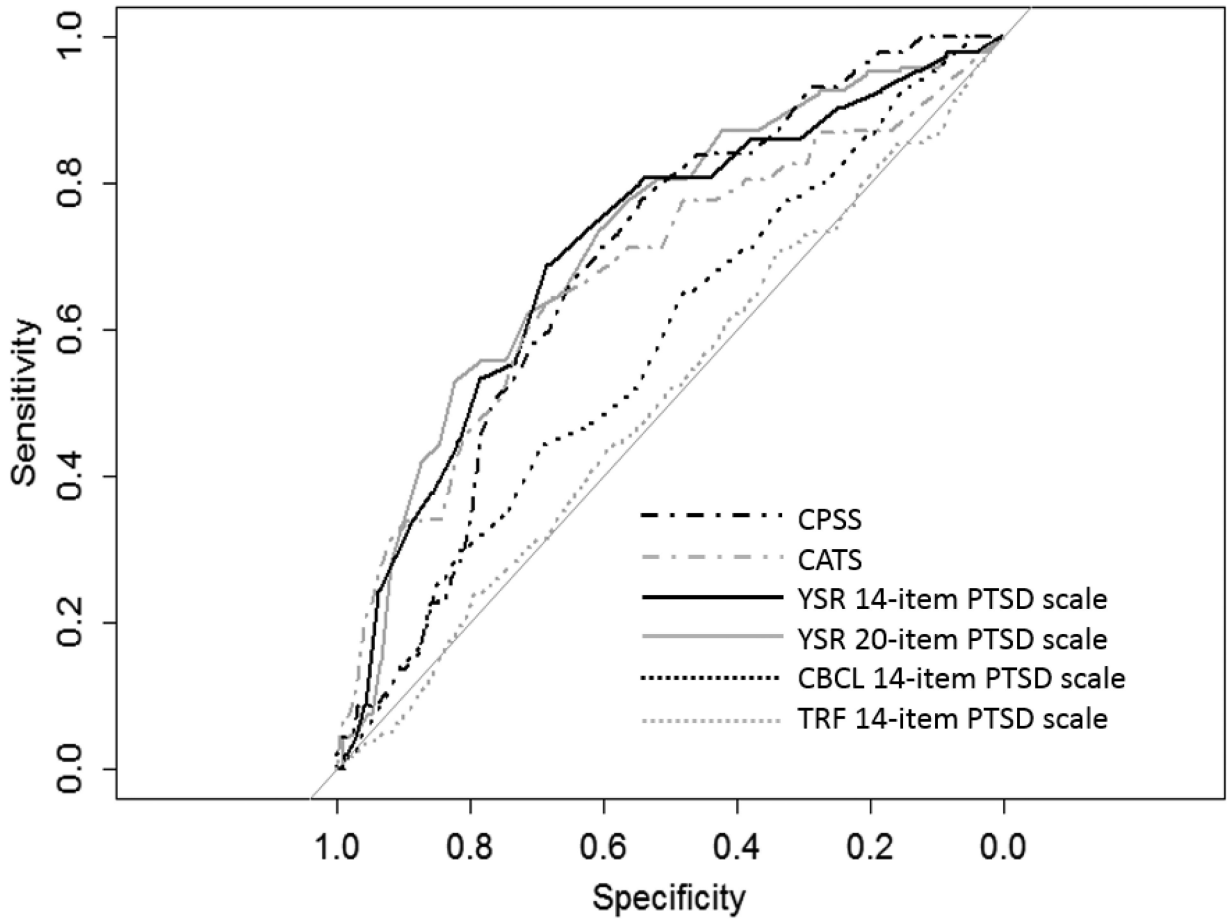


Figure 2.
The Receiver Operating Characteristic (ROC) curves for index tests predicting KSADS PTSD diagnoses ($N = 458$)
Note: YSR-14 item PTSD scale sums items 3, 8, 9, 11, 31, 34, 45, 47, 50, 52, 69, 87, 103, and 111.
YSR-20 item PTSD scale sums items on YSR-14 PTSD scale minus item 31 and plus items 29, 56b, 56c, 56f, 56g, 86, and 100.

Table 1

Diagnostic Characteristics (N = 458)

Characteristic	n	%
KSADS total sample		
Disruptive behavior (CD, DBD, ODD)	292	64
Any ADHD disorders	235	51
Unipolar disorder (major depressive disorder, dysthymia, NOS)	155	34
Anxiety disorder (acute Stress, GAD, phobia, separation anxiety)	107	23
Bipolar disorder (Type I, II, NOS, cyclothymia)	90	20
Adjustment disorders	67	15
PTSD	38	8
Any substance abuse or dependence	27	6
Psychotic disorder (schizophrenia, NOS)	18	4
Other Axis I disorders	89	19
KSADS reference standard positive: PTSD (regardless of comorbidity)		
PTSD, n (% of 38)	38	
Without comorbid	3	8
With comorbid current disorder		
Disruptive behavior (CD, DBD, ODD)	22	58
Unipolar disorder (major depressive disorder, dysthymia, NOS)	19	50
Anxiety disorder (acute Stress, GAD, phobia, separation anxiety)	16	42
Any ADHD disorders	15	39
Adjustment disorders	5	13
Bipolar disorder (Type I, II, NOS, cyclothymia)	4	11
Psychotic disorder (schizophrenia, NOS)	3	8
Other Axis I (enuresis)	1	3
KSADS reference standard negative: No PTSD (regardless of other diagnoses)		
No-PTSD, n (% of 420)	420	
Disruptive behavior (CD, DBD, ODD)	270	64
Any ADHD disorders	220	52
Unipolar disorder (major depressive disorder, dysthymia, NOS)	136	32
Anxiety disorder (GAD, OCD, panic disorder, phobia, separation anxiety, NOS)	91	22
Bipolar disorder (Type I, II, NOS, cyclothymia)	86	20
Adjustment disorders	62	15
Any substance abuse or dependence	27	6
Psychotic disorder (schizophrenia, NOS)	15	4
Other Axis I disorders *	88	21
No diagnosis	12	3

Note. An individual could have more than one comorbid disorder. Mean number of comorbid disorders was 2.4 ($SD = 1.3$, mode = 3) for the PTSD and 1.5 ($SD = 1.2$, mode = 1) for the no-PTSD groups.

* anorexia nervosa, bereavement, enuresis, encopresis, intermittent explosive disorder, any learning or pervasive developmental disorders, mood disorder NOS, night terrors, substance induced mood disorder or psychosis, and tic disorders.

Table 2

Index Test Correlations and Comparison of PTSD Symptom Severity Scores between Youths with and without PTSD Diagnoses (N=458)

Correlations	CPSS	CPSS+Impair	CATS	YSR-14	YSR-20	YSR-Int	CBCL-14	CBCL-20	CBCL-Int	TRF-14	TRF-20	TRF-Int
<i>Youth Report</i>												
CPSS	1.00											
CPSS + impairment	.98***	1.00										
CATS	.60***	.61***	1.00									
YSR-14	.61***	.61***	.60***	1.00								
YSR-20	.62***	.62***	.58***	.96***	1.00							
YSR Internalizing T score	.60***	.60***	.56***	.89***	.91***	1.00						
<i>Caregiver Report</i>												
CBCL-14	.16**	.16***	.13**	.24**	.22**	.21**	1.00					
CBCL-20	.17**	.17***	.14*	.26**	.26**	.24**	.94***	1.00				
CBCL Internalizing T score	.20**	.20***	.16**	.26**	.26**	.28**	.80***	.84***	1.00			
<i>Teacher report</i>												
TRF-14	.12	.08	-.10	-.02	-.06	-.04	.08	.03	.07	1.00		
TRF-20	.12	.08	-.02	.01	-.02	-.01	.12**	.06	.10*	.95***	1.00	
TRF Internalizing T score	.19	.16	-.06	.09	.08	.08	.04	.04	.14**	.74***	.75***	1.00
AUCs												
	.70***	.70***	.69***	.71***	.73***	.66***	.61*	.61*	.64**	.47	.44	.41
SE												
	.04	.04	.05	.05	.04	.05	.05	.05	.04	.08	.08	.08
95% CI(Lower/Upper)												
	.62/.78	.52/.67	.62/.78	.62/.80	.64/.81	.57/.76	.51/.71	.52/.71	.55/.73	.32/.62	.28/.59	.26/.57
Cohen's d for AUC												
	.74	.74	.70	.78	.87	.59	.40	.40	.50	-	-	-
No PTSD (n = 420) M(SD)												
	13.5 (11.4)	16.7 (11.7)	12.0 (8.8)	9.01 (5.9)	12.6 (7.9)	56.3 (12.9)	11.0 (5.2)	14.3 (6.6)	63.7 (10.2)	7.0 (4.6)	9.6 (6.2)	58.7 (10.0)
PTSD (n = 38) M(SD)												
	21.0 (10.9)	25.1 (11.4)	18.6 (10.2)	13.3 (5.8)	18.3 (7.1)	62.7 (12.4)	12.9 (5.2)	16.9 (6.4)	68.6 (7.9)	6.4 (4.7)	8.1 (6.4)	55.6 (9.5)
Cohen' d for PTSD												
	0.67	0.73	0.69	0.73	0.76	0.51	0.37	0.40	0.54	-0.13	-0.24	-0.32

Note.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

The p values of AUCs based on bootstrapping 2000 resamples with replacement per data file and adjusted for variance due to imputation. Cohen's d for AUC were from Rice and Harris's reference table (2005)

* $p < .05$

** $p < .01$

*** $p < .001$, two tailed.

Table 3
 Diagnostic accuracy statistics associated with different optimal threshold and multi-level diagnostic likelihood ratios

Cut Score	Sensitivity	Specificity	Kappa	Level	DLR+	DLR-	Prevalence of 10%			Prevalence of 40%		
							PPV	NPV	DLR	PPV	NPV	PPV
YSR-14												
90% Sensitivity: 4+	.92	.18	.02	.71	1.12	0.45	.11	.95	.43	.77		
90% Specificity*: 17+	.37	.89	.21	.13	3.36	0.71	.27	.93	.69	.68		
<i>Multilevel Diagnostic Likelihood Ratios (DLRs) (based on informative thresholds)</i>												
0 to 10 (lower risk)	--	--	--	(~60%)	0.47	--	.05	--	.24	--		
11 to 17 (neutral)	--	--	--	(~30%)	1.35	--	.13	--	.47	--		
18+ (higher risk)	--	--	--	(~10%)	3.64	--	.29	--	.71	--		
CPSS												
90% Sensitivity: 6+	.92	.28	.04	.74	1.27	0.30	.12	.97	.46	.83		
Max. Kappa: 21+	.56	.77	.16	.26	2.37	0.58	.21	.94	.61	.72		
90% Specificity: 30+	.19	.88	.06	.13	1.59	0.92	.15	.91	.52	.62		
<i>Multilevel Diagnostic Likelihood Ratios (DLRs) (based on informative thresholds)</i>												
0 to 12 (lower risk)	--	--	--	(~50%)	0.43	--	.05	--	.22	--		
13 to 20 (neutral)	--	--	--	(~25%)	0.89	--	.09	--	.37	--		
21+ (higher risk)	--	--	--	(~25%)	2.37	--	.21	--	.61	--		

Note. Sensitivity=portion of cases with PTSD diagnoses correctly identified; specificity=portion of cases without PTSD correctly identified; kappa=accuracy above chance performance; level=portion of sample testing above threshold; DLR+ is diagnostic likelihood ratio associated with a positive test, DLR- is diagnostic likelihood ratio associated with a negative test result; PPV=positive predictive value, or the portion of test positives that have PTSD; NPV=negative predictive value, or the portion of test negatives that do not have PTSD. The prevalence rates of 10% and 40% represent lower and upper bounds of what typically would be encountered in outpatient mental health,(Gabbay et al., 2004) providing a sense of how the instruments might work in practice.