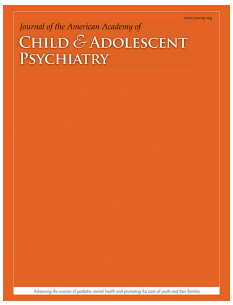
Accepted Manuscript



Accuracy of Achenbach Scales in the Screening of Attention-Deficit/Hyperactivity Disorder in a Community Mental Health Clinic

Joseph S. Raiker, PhD, Andrew J. Freeman, PhD, Guillermo Perez-Algorta, PhD, Thomas W. Frazier, PhD, Robert L. Findling, MD, MBA, Eric A. Youngstrom, PhD

PII: S0890-8567(17)30071-0

DOI: 10.1016/j.jaac.2017.02.007

Reference: JAAC 1717

To appear in: *Journal of the American Academy of Child & Adolescent Psychiatry*

Received Date: 20 August 2016

Revised Date: 19 February 2017

Accepted Date: 21 February 2017

Please cite this article as: Raiker JS, Freeman AJ, Perez-Algorta G, Frazier TW, Findling RL, Youngstrom EA, Accuracy of Achenbach Scales in the Screening of Attention-Deficit/Hyperactivity Disorder in a Community Mental Health Clinic, *Journal of the American Academy of Child & Adolescent Psychiatry* (2017), doi: 10.1016/j.jaac.2017.02.007.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Accuracy of Achenbach Scales in the Screening of Attention-Deficit/Hyperactivity Disorder in a Community Mental Health Clinic

Joseph S. Raiker, PhD, Andrew J. Freeman, PhD, Guillermo Perez-Algorta, PhD, Thomas W. Frazier, PhD, Robert L. Findling, MD, MBA, Eric A. Youngstrom, PhD

This article is discussed in an editorial by Dr. John Hamilton on page xx.

Clinical guidance is available at the end of this article.

Accepted February 27, 2017

Dr. Raiker is with Florida International University, Miami. Dr. Freeman is with the University of Nevada, Las Vegas. Dr. Perez-Algorta is with Lancaster University, Lancashire, UK. Dr. Frazier is with the Center for Autism at Cleveland Clinic Lerner College of Medicine, Cleveland. Dr. Findling is with Johns Hopkins University, Baltimore. Dr. Youngstrom is with the University of North Carolina at Chapel Hill.

The work was supported in part by National Institute of Mental Health Grant NIHR01 MH066647 (Principal Investigator, Eric A. Youngstrom).

Drs. Freeman and Youngstrom served as the statistical experts for this research.

The authors thank the families who participated in this research.

Disclosure: Dr. Raiker has received research support from the Brain and Behavior Research Foundation, the Children's Trust, NIMH, and NSF. Dr. Frazier has received research support from, acted as a consultant for, received travel support and/or speaker's honorarium from the Cole Family Research Fund, Simons Foundation, Ingalls Foundation, Forest Laboratories, Ecoeos, IntegraGen, Kugona LLC, Shire Development, Ohio Third Frontier, Bristol-Myers Squibb, NIH, and the Brain and Behavior Research Foundation. Dr. Findling has received research support from, acted as a consultant for, and/or served on a speaker's bureau for Akili, Alcobra, American Academy of Child and Adolescent Psychiatry, American Psychiatric Press, Bracket, Epharma Solutions, Forest, Genentech, Guilford Press, Ironshore, Johns Hopkins University Press, KemPharm, Lundbeck, Medgenics, Merck, NIH, Neurim, PCORI, Pfizer, Physicians Postgraduate Press, Purdue, Roche, Sage, Shire, Sunovion, Supernus Pharmaceuticals, Syneurx, Takeda, Teva, Tris, Validus, and WebMD. Dr. Youngstrom has received grant support from the NIMH, the Society for Clinical Child and Adolescent Psychology, the American Psychological Association, and the Association for Psychological Science. He has served as a consultant to Pearson Publishing, Joe Startup Technologies, Janssen, Lundbeck, and Western Psychological Services about psychological assessment. Drs. Freeman and Perez-Algorta report no biomedical financial interests or potential conflicts of interest.

ACCEPTED MANUSCRIPT

Correspondence to Andrew J. Freeman, PhD, University of Nevada, Las Vegas, MS 5030, 4505 S. Maryland Pkwy., Las Vegas, NV 89154-5030; email: andrew.freeman@unlv.edu.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ABSTRACT

Objective: Use receiver operating characteristics analysis to identify multilevel diagnostic likelihood ratios and provide a framework for the diagnosis of attention-deficit/hyperactivity disorder (ADHD) in children (5-10 years) and adolescents (11-18 years) in an outpatient setting.

Method: Parent, teacher, and youth reports from the Achenbach System of Empirically Based Assessment (ASEBA) were obtained for 299 children and 321 adolescents with multiple imputation of missing data. The reference standard was diagnosis of ADHD based on case history and a semi-structured diagnostic interview masked to the ASEBA measures.

Results: In children, caregiver-reported Attention Problems (area under the curve [AUC]=.74) outperformed all other subscales of the caregiver and teacher measures (AUCs≤.72). In the older sample, caregiver- and teacher-reported Attention Problems (parent AUC=.73; teacher AUC=.61) were best at identifying ADHD. Inclusion of parent- and teacher-report significantly (all *p*s <.001) increased prediction of ADHD diagnosis whereas youth self-report did not.

Conclusion: Parent-reported Attention Problems were more useful than teacher- and self-report in identifying ADHD. Combining parent and teacher report improved identification. Multilevel likelihood ratios are provided to facilitate routine clinical use.

Key Words: ADHD, children and adolescents, sensitivity and specificity, likelihood ratios, receiver operating characteristic curve

INTRODUCTION

Despite decades of research on the assessment of attention-deficit/hyperactivity disorder (ADHD), a single diagnostic test for the disorder remains elusive. Diagnosis is complicated by the lack of specificity for symptoms (e.g., inattention) that occurs across other forms of psychopathology (e.g., depression). Practice guidelines recommend a multi-informant and multi-

ACCEPTED MANUSCRIPT

method assessment with information obtained from multiple settings such as home and school for youth; however, little guidance is available regarding how to interpret information from multiple informants.[1] Information is typically collected via interviews with parent and child as well as the use of one or more parent and/or teacher rating scales; conventional guidance recommends careful consideration of their psychometric properties while also weighing their limitations.[2]

Pelham et al.[2] have highlighted the use of both narrowband (e.g., ADHD-specific rating scales) and broadband rating scales (e.g., Child Behavior Checklist [CBCL]) in the assessment of ADHD. Both types show adequate reliability, validity, and utility at different times throughout the assessment process. Broadband scales are most useful during the screening phase, as they assess an array of behavior and emotional difficulties associated with various forms of psychopathology (e.g., anxiety) and may help narrow the focus of subsequent assessment. Narrowband scales measure symptoms related to a specific disorder, strengthening confidence in a particular diagnosis once a candidate diagnosis has been identified during initial screening.[3]

The Achenbach Scales of Empirically Based Assessment (ASEBA; [4]—the Child Behavior Checklist (CBCL), Teacher Report Form (TRF), and Youth Self-Report Form (YSR)—are commonly used scales in children and adolescents. CBCL subscales differentiate youths with ADHD from youths without ADHD.[5-13] However, analyses usually group youth based on known diagnoses, and then test mean score differences between children with and without ADHD. In contrast, clinical decision-making typically reverses the order: clinicians obtain a score on a measure and then must determine the likelihood that the youth has ADHD.[3] Positive predictive power (PPP) and negative predictive power (NPP) attempt to mitigate this conflict and improve clinical decision making by providing estimates of the likelihood that an individual with a

ACCEPTED MANUSCRIPT

particular score has or does not have the disorder. Despite their improved clinical utility, these values change as a function of the prevalence of the condition. Diagnostic likelihood ratios (DLRs) provide an estimate of the likelihood that a given score indicates the presence (DLR+) or absence (DLR-) of a particular disorder and are not sensitive to prevalence. A nomogram allows for a priori estimates of the likelihood of a diagnosis (e.g., prevalence) to be combined with the DLR to create PPP and NPP.

Evidence-based medicine methods help with score interpretation and guiding clinical decision-making.[14] Clinicians combine the pretest probabilities of having a diagnosis (e.g., base rate) with diagnostic likelihood ratios (DLRs) derived from scores on the screening test results (e.g., [15,16]) using an inexpensive tool such as the nomogram (Figure 1). These interpretive methods produce large gains in consistency and accuracy.[17]

The ADHD base rate can be the pretest probability estimate. ADHD occurs in 3–7% of school-age children,[18] varying somewhat across sex,[19-21] age,[21] and ethnicity.[20] Rates of ADHD are substantially higher in outpatient clinic-based samples, with estimates ranging from 23 to 58%.[22,23] If the base rate of ADHD in a clinic is known (e.g., electronic medical record), then clinicians could begin with their clinic base rate. Otherwise, clinicians could use base rates from similar clinics.

Next, the DLR of a youth's score on a measure revises the probability that a youth with this score has ADHD. DLRs ranging from one to infinity increase the likelihood of a diagnosis, whereas DLRs ranging from 0 to 1 decrease the likelihood of a diagnosis. A DLR of 1 indicates no change in a youth's risk for ADHD. CBCL *T*-scores between 50 and 75 have been associated with DLRs ranging from .99 to 34 in community, school, and clinic settings[6,8-10] suggesting that the CBCL's ability to discriminate between children with and without a diagnosis of ADHD

ACCEPTED MANUSCRIPT

varies depending upon the clinical setting and cut-score selected. Despite its widespread use and practice parameters calling for the integration of multiple informants, less information is available concerning DLRs based on scores from the ASEBA scales completed by teacherscf. [11] and adolescent self-report. Adolescent self-report has likely been excluded from prior work due to samples that focus primarily on children and decades of work suggesting that self-report of ADHD is poor.cf. [24,25-28] Furthermore, most studies examining CBCL diagnostic efficiency have compared children with ADHD to healthy children without clinical diagnoses[6,9,10] instead of children with other psychiatric diagnoses.cf. [5,13] In most clinical decision-making contexts, healthy controls are not an informative comparison. Rarely is the clinical question, Does this child have ADHD or no diagnosis? Instead, the question is usually, Does this child have ADHD, some other diagnosis, or comorbid diagnoses? Although past work has included both children and adolescents, diagnostic efficiency and DLRs have not been examined separately for these two age groups despite unique diagnostic challenges inherent to the diagnosis of ADHD in adolescence, e.g., [29], nor across caregiver, youth, and teacher report in the same sample.

This study is the first to use receiver operating characteristics (ROC) and multilevel DLRs while capitalizing on the full range of scores to provide estimates of diagnostic efficiency across ASEBA scales. Specifically, to aid clinicians in clinical diagnosis, ROC will be employed to create multilevel DLRs for the CBCL, TRF, and YSR in youth in a clinical sample that can then be used to aid diagnostic decision-making. We expect diagnostic efficiency to be lower relative to previous investigations that included healthy controls.[5,6,9,10,13] We expect both parent and teacher report to show incremental validity predicting ADHD status.[11]

METHOD

Participants

ACCEPTED MANUSCRIPT

Participants (5 to 18 years old) were recruited using a prospective, consecutive case series design from all intakes at an urban, community mental health center between July 2003 and March 2008 regardless of presenting reason. Inclusion criteria were: (a) both caregiver and youth presented for the assessment and (b) both were conversant in English. The institutional review board at University Hospitals of Cleveland approved the procedures. All caregivers provided written informed consent, and all youth provided assent.

Measures

Diagnosis. Assessments were completed using the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) – Present and Lifetime version.[30] Training required that research assistants provide passing ratings on five interviews led by trained raters, followed by administering five interviews while being observed by a trained rater. Highly trained raters passed by achieving an overall $\kappa \geq .85$ at the symptom level and $\kappa = 1.0$ at the diagnosis level.

A clinical psychologist assigned diagnoses using the longitudinal evaluation of all available data (LEAD) standard[31] after reviewing: (a) the diagnostic interview, (b) clinical intake, and (c) all other available information (e.g., school records, treatment history). Both research assistant and psychologist were blind to the parent-, self-, and teacher-report questionnaires. Diagnoses of ADHD were made in accordance with *DSM-IV-TR*.[32]

Index Tests

Achenbach System of Empirically Based Assessment (ASEBA; Achenbach, 2001). The ASEBA includes the Child Behavior Checklist (CBCL), Teacher Report Form (TRF), and Youth Self-Report (YSR). Each measure contains 118 problem behavior items rated 0 (not at all typical of the child) to 2 (often typical of the child). Caregivers and teachers completed the CBCL or TRF 6-18 years. Caregivers and teachers of youth aged 5 completed the 1.5–5.5-year version.

ACCEPTED MANUSCRIPT

Analyses used the empirically-derived subscales of Attention Problems, Externalizing Problems, and the *DSM*-oriented ADHD subscale, as these have the most relevant content and performed best in prior work. The *DSM*-oriented ADHD subscale was constructed by experts identifying the seven items most consistent with *DSM*-defined ADHD and shares fives items with the Attention Problems subscale (10 items). The Venkatraman difference test accounts for correlated measures in the ROC analyses.

Procedure

Research assistants met with the caregiver and youth individually and sequentially to conduct the semi-structured interview (additional details provided in [33]) and a separate research assistant gathered the questionnaires. A release of information form was obtained, and questionnaires were mailed directly to the youth's teacher.

Statistical Methods. All participants completed the reference standard (KSADS). Index tests (CBCL and YSR) were completed by 98% and 96% of children and adolescents, respectively. Missing data was attributed primarily to the TRF (36% overall return rate). Multiple imputation (m = 10) was conducted after verifying that the influence of missing data was negligible (largest r_{pb} = .11, p = .07), and there were no significant patterns of missingness via the MICE package in R.[34] Briefly, multiple imputation involves generating values for missing data by utilizing the available information from collected data as predictors. This process is repeated a predetermined number of times (denoted as m) until stable estimates for the generated values are obtained.

Methods for calculating and comparing diagnostic accuracy. Youth with all subtypes of ADHD were compared to all other youth regardless of other *DSM-IV-TR* Axis I diagnoses using ROC curves. The area under the curve (AUROC) represents the diagnostic efficiency of the

ACCEPTED MANUSCRIPT

measure. An AUROC of .50 indicates the measure performs at chance levels. An AUROC of 1.0 indicates the measure performs perfectly. The following AUROC benchmarks have been suggested by multiple sources: ≥ .90 are "excellent," ≥ .80 are "good," ≥ .70 are "fair," and ≤ .70 are "poor" [35]; however, AUROCs of .7 - .8 are considered realistic of a good test.[16] Specific subscales of the CBCL, TRF, and YSR were compared both within and across informants using Venkatraman's test that compares the area between the related ROC curves.[36,37] All ROC analyses were performed using pROC in R.[38] Logistic regression examined whether combinations of measures from the same rater or across raters provided incremental utility. Finally, multilevel DLRs provided interpretative guidance for integrating the evidence-based medicine approach (described above) into the diagnosis of ADHD in clinical practice.[39] DLRs are estimated by obtaining ratios of the number of true positives (sensitivity) to false positives (1-specificity) and false negatives (1-sensitivity) to true negatives (specificity) to obtain positive (DLR+) and negative (DLR-) DLRs, respectfully. DLRs range from 0 to positive infinity. A DLR greater than 1 indicates the result is associated with a greater likelihood of having a diagnosis of ADHD, and a DLR less than 1 indicates the result is associated with a decreased likelihood of having a diagnosis of ADHD.

RESULTS

Participants

Children (n = 299, age 5-11) and adolescents (n = 321, age 11-18) were split into two groups. Children were significantly more likely to have ADHD, DLR+ = 1.98, than adolescents, DLR- = .59, $\chi^2(1)=46.92$, p<.0001. Males were significantly more likely to have ADHD in both children, DLR+ = 1.69, DLR- = .46, $\chi^2(1)=21.06$, p<.0001, and adolescents, DLR+= 1.84, DLR- = .50, $\chi^2(1)=31.34$, p<.0001. Adolescents with ADHD (M=12.99, SD=1.71) were significantly

ACCEPTED MANUSCRIPT

younger than adolescents without ADHD (*M*=13.91, *SD*=1.89), *t*(307.53)=4.56, p<.0001. No race differences were observed between groups in either age group (Table 1).

Diagnostic Efficiency

Caregiver-report measures demonstrated large effect sizes (Table 2). In contrast, teacher-report measures demonstrated small to moderate effect sizes, and youth self-report measures demonstrated small effect sizes when comparing youth with and without ADHD (Table 2). AUROC values (Figure 2) indicated that parent-report subscales were "fair" and clinically useful; teacher-report was "poor" but could be clinically useful; and youth self-report was "poor" and not clinically useful.

In children, all caregiver-reported CBCL subscales and teacher-reported Externalizing and ADHD Problems performed significantly better than the teacher-reported Attention Problems subscale, *p*s < .05. There were no significant differences among the caregiver-reported subscales of the CBCL, *p*s > .10. Teacher-reported Externalizing was significantly better than teacher-reported ADHD Problems, *p* < .05. In adolescents, the caregiver-reported CBCL subscales performed significantly better than teacher-report or youth self-report, *p*s < .05. Teacher and youth self-report were not significantly different, *p*s > .10. Within informant, subscales were typically not significantly different unless otherwise noted, *p*s > .10. Caregiver-report of adolescent symptoms was not significantly different from caregiver-report of child symptoms, *p*s > .05. Teacher-report of adolescent symptoms was not significantly different from teacher-report of child symptoms, *p*s > .05.

Combinations of Index Tests

The caregiver-reported Attention Problems subscale had the strongest diagnostic accuracy across both ages. Therefore, logistic regression evaluated whether including different

ACCEPTED MANUSCRIPT

subscales from the same rater (caregiver-reported Externalizing or ADHD subscales) or subscales from other informants (e.g., CBCL Attention Problems and TRF Attention Problems) significantly improved prediction above the caregiver-reported Attention Problems subscale alone. The incremental utility of an additional score and the interaction term that evaluates the combination of the measures were examined utilizing hierarchical logistic regression. Similar to the ROC analyses, the CBCL Attention Problems subscale significantly predicted ADHD in children and adolescents (Cox and Snell R^2 = .11, .14; *p*s < .001).

Adding either the CBCL Externalizing (ΔR^2 = .02 in children, .05 in adolescents) or the CBCL ADHD subscales (ΔR^2 = .04 in children, .07 in adolescents) resulted in an increase in incremental utility, *p*s < .01. Interaction terms were significant only in children for both the CBCL Externalizing (ΔR^2=.04 in children, .01 in adolescents) and the CBCL ADHD subscales (ΔR^2=.03 in children, <.01 in adolescents). The interaction indicated among parent-reported subscales that if one score is high and one score is low, to interpret the high score among the parent-report scales.

Adding teacher-report to the CBCL Attention Problems subscale resulted in incremental improvements in prediction of ADHD. For children, adding the teacher-report of Externalizing Problems (ΔR^2= .04, *p*<.01) and ADHD Problems (ΔR^2= .03, *p*<.01) resulted in an incremental improvement in diagnostic efficiency, but adding the teacher-reported Attention Problems subscale (ΔR^2= .00, *p*>.10) did not. None of the interaction terms between the parent-reported Attention Problems subscale and the teacher-report subscales were significant for children, ΔR^2<= .01, all *p*s >.05. Among adolescents, including the teacher-reported Attention Problems subscale (ΔR^2= .07, *p*<.01) improved incremental utility, but the Externalizing (ΔR^2= .01, *p*>.10) and ADHD Problems subscales (ΔR^2=.01, *p*>.10) did not. However, both the teacher-reported

ACCEPTED MANUSCRIPT

Externalizing and ADHD Problems subscales interacted with the parent-reported Attention Problems subscales such that low scores on the teacher scales do not negate ADHD risk whereas high scores on the combinations increase ADHD risk.

Including youth self-report scales did not significantly improve classification after controlling for caregiver-reported Attention Problems, all *p*s > .10. Collectively, inclusion of additional informants and/or subscales beyond the parent-reported Attention Problems subscale resulted in slight increases in the overall prediction accuracy. Table 3 presents the diagnostic likelihood ratios for subscales by informant.

DISCUSSION

Although broadband rating scales completed by parents and teachers differentiate youth with ADHD from youth without ADHD,[5-13] applying these findings to clinical settings is limited by a number of factors. First, practitioners must determine the likelihood of a diagnosis by examining test results (e.g., percentiles), whereas most research in this area is based on how well those test results predict an already known diagnosis (e.g., based on a semi-structured interview), which is of limited clinical value. Additionally, prior research has relied on comparing youth with ADHD to youth without, the results of which answer the question of whether this child has ADHD or is a healthy child (for exceptions, see [5,13]). This comparison is artificial given that clinicians are usually faced with a decision regarding whether the child has ADHD, some other diagnosis, or multiple diagnoses. This study sought to extend previous findings regarding the utility of parent, teacher, and youth self-report in diagnosing ADHD in a clinical sample using ROC. Additionally, this is the first study to provide clinically useful multilevel DLRs to aid clinicians in applying an evidence-based medicine approach to the diagnosis of ADHD in their own clinics.

ACCEPTED MANUSCRIPT

The CBCL and TRF Attention Problems subscales demonstrated better utility than general scales such as the Externalizing Problems subscale in predicting a diagnosis of ADHD consistent with past findings.[6,10] Additionally, parent-report of Attention Problems was a better predictor of ADHD than teacher-report, particularly in younger children, despite past reports of greater predictive utility from teacher report,[11] a discrepancy that may be attributable to differences in setting as well as diagnoses in the non-ADHD comparison group. Specifically, past findings were based on samples recruited from research clinics targeting children with potential ADHD symptoms,[11] whereas the current sample includes a broader range of referrals given the use of a community mental health center. Prior reports included a greater proportion of children with internalizing disorders in the non-ADHD comparison group, whereas the current sample of non-ADHD youth included children with disorders that may contain features that are behaviorally more similar to ADHD (e.g., bipolar disorders, psychotic disorders) resulting in teachers experiencing greater difficulty discriminating between ADHD and non-ADHD. As expected, youth self-report of attention difficulties did not discriminate youth with ADHD from youth without ADHD, consistent with past findings (e.g., [29]). Collectively, our findings are consistent with work indicating that specific ADHD symptoms are better than general externalizing symptoms for diagnostic accuracy of ADHD (e.g., [6,40]). While some have argued that teacher-report is biased toward labeling negative behavior as attention problems,[41] our findings indicate that overall teacher-report demonstrated low sensitivity and high specificity, suggesting that teachers were missing most cases of ADHD but were accurate when they did identify ADHD.

For all ages, diagnostic accuracy is somewhat enhanced when parent- and teacher-report are used in combination. However, the incremental utility of teacher-report information was

ACCEPTED MANUSCRIPT

negligible, and parent- and teacher-report were weakly associated, consistent with prior work [11,23,42] indicating that the information provided by teachers and parents is largely overlapping and that adding teacher-report provides only a slight increase in accuracy once parent-report is considered.

Diagnosing ADHD accurately provides the bedrock for efficacious and targeted intervention. The evidence-based assessment approach described above can be combined with the results of the current study in a theoretical case such as a 7-year-old referred for treatment with a parent-reported Attention Problems *T*-score of 75 and a teacher-reported Attention Problems *T*-score of 70 by using a nomogram (Figure 2). In the current sample, the base rate of ADHD for children (78%) is placed on the left axis of the nomogram. The DLR for a Clinical Score (1.97) on caregiver-reported Attention Problems is placed on the middle axis. A line connecting the two numbers provides an updated posterior probability (87%). If teacher-reported Attention Problems is added, the posterior probability (87%) from the prior step becomes the base rate and is placed on the left axis. The DLR for a clinical score from the teacher (1.31) is placed on the middle axis. For every 100 children with this set of scores, approximately 90 will meet criteria for ADHD in a community mental health clinic. Overall, using Bayesian approaches when screening for a common clinical diagnosis such as ADHD can help direct finite clinical (e.g., referral for neuropsychological testing, behavior therapy) and educational (e.g., tutors) resources. For another example, see Figure 3.

Strengths of the present study include: 1) adherence to the Standards for Reporting Diagnostic Accuracy Studies (STARD) guidelines for reporting diagnostic test results,[43] 2) large samples in both age groups and evaluation of the diagnostic efficiency of these scales in these age groups separately, 3) ADHD diagnoses masked to the ASEBA results, 4) examination of

ACCEPTED MANUSCRIPT

parent, teacher, and youth report in the same sample, and 5) use of multiple methods for evaluating diagnostic efficiency (i.e., global estimates, multi-level DLRs), which provide a clinically meaningful way of interpreting test scores for practitioners. The primary limitation of the present study was the diagnosis of ADHD being based on information available at the time of assessment (i.e., parent interview, youth interview, behavioral observations, and review of records) and not incorporating teacher reports. While methods might bias findings toward improved diagnostic efficiency for parent and youth report, our data indicate that only parent and teacher report were predictors of ADHD diagnoses, and our findings are consistent with recent unblended consideration of parent, teacher, and youth self-report.[44] Additionally, multiple imputation was performed to produce unbiased teacher-report estimates, avoiding a potential source of bias in test evaluation.[43] Our procedures likely mimic best-case clinical practice in which parent and children are interviewed separately and teacher report is obtained post hoc, if at all. Diagnostic efficiency estimates of parent and teacher ASEBA scales fall within the "useful" but not "high" ranges of discrimination,[35] consistent with previous studies comparing individuals with ADHD to other clinical conditions.[45] This finding emphasizes the need for appropriate comparison groups when evaluating test performance. Future work should compare ASEBA data to *DSM*-based narrowband scales.[4] Diagnostic efficiency of *DSM*-based narrowband scales might show greater discrimination, although, as mentioned previously, these scales may be more susceptible to informant biases.[46] Finally, while the high base rate of ADHD in the current sample was in the optimal range for Bayesian decision-making, the DLRs will result in different assessments of risk when applied to low base rate settings.[47] Clinicians need to determine whether their practices are similar enough in diagnostic caseload to our sample; otherwise, the DLRs are likely to be inaccurate.

ACCEPTED MANUSCRIPT

Collectively, the current study replicates and extends previous findings that parent and teacher report of behavioral problems discriminate between children with and without a diagnosis of ADHD, even in settings where a broader range of psychiatric disorders is likely to be observed. The risk of a youth with "clinical" range scores from caregivers or teachers on Attention Problems increases by approximately 15%, whereas "normal" range scores reduce risk by 25-30%. Additionally, incorporating youth self-report of behavioral problems is unlikely to improve diagnostic decision-making and combining parent and teacher report results in small improvements in diagnostic efficiency. This is the first study to provide clinicians with multilevel DLRs that can be applied to their own practice using an evidence-based medicine approach that incorporates low-cost tools (e.g., nomogram). Finally, it is of crucial importance to note that no combination of scores resulted in 100% accuracy, and questionnaires are not intended to be diagnostic, as they do not systematically assess all relevant clinical features of a disorder (e.g., onset, duration, course, or impairment). In short, questionnaires provide a cost-effective and efficient approach to screen for disorders and helping clinicians prioritize more expensive diagnostic procedures. Questionnaire usefulness is improved drastically when combining DLRs based on scores from these questionnaires with a priori estimates of the likelihood of having a diagnosis of ADHD (e.g., base rate). Future work investigating the incremental utility of incorporating additional methods of assessment (e.g., neurocognitive testing, genetic testing, neuroimaging) is warranted.

Clinical Guidance

- Parents and teachers often provide discrepant accounts of youth's problem behavior when completing rating scales related to ADHD behaviors; however, recommendations concerning the integration of information from multiple informants in the assessment of ADHD are lacking.

ACCEPTED MANUSCRIPT

- In an outpatient community, mental health setting, caregiver, youth, and teacher report predict whether youth meet criteria for ADHD.
- Clinical range scores (*T*-score > 70) from caregivers or teachers double the odds of a youth meeting criteria for ADHD, and caregiver reports in normal range (*T*-score < 64) decrease the likelihood that a youth will meet criteria for ADHD. Youth self-report does not substantially inform ADHD decision-making.
- Combining caregiver and teacher reports changes a youth's odds for ADHD mildly. When information is available from both caregivers and teachers, clinicians should weight the more severe report more strongly.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. Demographic and Sample Characteristics

Characteristic	Age 5 to 11 (*n*=299)	Age 11 to 18 (*n*=321)
Age in years (SD)	7.63 (1.65)	13.43 (1.85)
Gender (Male)	202 (68%)	172 (54%)
Ethnicity		
African-American	260 (87%)	287 (89%)
Hispanic	8 (3%)	0 (0%)
White	19 (6%)	20 (6%)
Other	12 (4%)	14 (4%)
Any ADHD (regardless of comorbidity)	235 (79%)	168 (52%)
ADHD Inattentive	28 (9%)	33 (10%)
ADHD Hyperactive Impulsive	26 (9%)	12 (4%)
ADHD Combined	159 (53%)	80 (25%)
ADHD NOS	22 (7%)	43 (13%)
Comorbid Axis 1 Diagnoses	2.82 (1.22)	3.26 (1.31)
Non-ADHD clinical comparison		
Bipolar disorder (BP-I, -II, -NOS, cyclothymia)	4 (1%)	16 (5%)
Unipolar depression (MDD or dysthymia)	16 (5%)	73 (23%)
Other disruptive behavior	14 (5%)	35 (11%)
Residual[a]	30 (10%)	29 (9%)
Comorbid axis 1 diagnoses	1.25 (1.12)	2.18 (1.36)

Note: Youth with and without attention-deficit/hyperactivity disorder (ADHD) diagnoses also met criteria for 1 to 8 (median = 3) other *DSM-IV* Axis I diagnoses. Adolescents had more comorbid diagnoses than children. Youth with ADHD had more comorbid diagnoses than youth without ADHD. BP-I, -II, -NOS = bipolar I, II, not otherwise specified; MDD = major depressive disorder.

[a]Anxiety, posttraumatic stress disorder, psychotic disorders, or no Axis I.

*p*s < .05.

ACCEPTED MANUSCRIPT

Table 2. Index Test Distributions for Youths With and Without Attention-Deficit/Hyperactivity Disorder (ADHD) Diagnoses

Age 5 to 11 (*n* = 299)									
		No ADHD (*n* =64)		ADHD (*n* =235)					
Informant	Index Test	*M*	*SD*	*M*	*SD*	AUROC (95% *CI*)	Cohen's *d*	*t*	*p*
Caregiver	Attention Problems	64.25	12.92	73.30	10.57	.74 (.66 - .82)	.88	5.78	<.001
	Externalizing	65.14	13.56	72.96	8.00	.68 (.60 - .77)	.67	4.41	<.001
	ADHD	63.37	10.22	71.35	7.97	.72 (.65 - .80)	.81	5.66	<.001
Teacher	Attention Problems	65.57	11.45	66.78	9.90	.56 (.47 - .65)	.21	.82	>.40
	Externalizing	62.73	8.14	67.80	9.19	.67 (.59 - .74)	.62	4.06	<.001
	ADHD	61.73	8.69	66.26	8.15	.62 (.55 - .70)	.43	3.67	<.001
Age 12 to 18 (*m* = 321)									
		No ADHD (*n* =153)		ADHD (*n* =168)					
Informant	Index Test	*M*	*SD*	*M*	*SD*	AUROC (95% *CI*)	Cohen's *d*	*t*	*p*
Caregiver	Attention Problems	64.16	11.15	73.15	11.39	.73 (.68 - .79)	.87	7.16	<.001
	Externalizing	64.45	10.07	72.97	7.33	.73 (.67 - .78)	.87	7.73	<.001
	ADHD	63.91	9.25	71.57	7.33	.73 (.67 - .78)	.87	8.15	<.001
Teacher	Attention Problems	62.35	9.22	65.47	9.88	.61 (.54 - .68)	.40	2.96	<.01
	Externalizing	62.02	10.51	64.61	9.60	.57 (.50 - .63)	.25	2.98	<.01
	ADHD	61.85	9.20	62.93	8.76	.56 (.50 - .62)	.21	2.39	.02
Youth	Attention Problems	62.35	9.22	65.47	9.88	.59 (.53 - .65)	.32	2.66	<.01
	Externalizing	56.73	11.56	59.85	11.37	.58 (.53 - .64)	.29	2.68	<.01
	ADHD	58.08	8.26	59.26	8.08	.56 (.49 - .62)	.21	2.44	.02

Note: Cohen's *d* of .3 = small, .5 = medium, and .8 = large effect size for the social sciences. Data reflect *T*-scores. AUROC = Area Under the Curve.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3. Multilevel Likelihood Ratios for Index Test Scores

Age 5-11 Likelihood Ratios – 78% prevalence of any ADHD					
		Range:	Normal Range	Borderline	Clinical
Informant	Measure	Score:	<64	64-69	≥70
Caregiver	Attention Problems*		.23	1.86	1.97
	Externalizing*		.23	1.35	1.67
	ADHD*		.23	1.15	1.67
Teacher	Attention Problems		.88	1.06	1.31
	Externalizing*		.58	1.22	2.24
	ADHD*		.58	1.15	2.24
Age 12-18 Likelihood Ratios – 52% prevalence of any ADHD					
		Range:	Normal Range	Borderline	Clinical
Informant	Measure	Score:	<64	64-69	≥ 70
Caregiver	Attention Problems*		.34	1.41	2.22
	Externalizing*		.31	.76	2.02
	ADHD*		.31	1.01	2.02
Teacher	Attention Problems*		.73	1.21	1.67
	Externalizing*		.83	1.14	1.33
	ADHD*		.83	1.25	1.33
Youth	Attention Problems*		.86	1.28	1.74
	Externalizing*		.83	1.28	1.61
	ADHD		.83	.76	1.61

Note: Ranges are based on Achenbach's recommended empirical interpretations. ADHD = attention-deficit/hyperactivity disorder.

* Receiver operating characteristics $p < .05$

ACCEPTED MANUSCRIPT

Figure 1. Probability nomogram.
Figure 2. Receiver operating characteristics for the most optimal subtests by informant and age range.
Figure 3. Case example of nomogram. Note: John Doe's parents bring him to the clinic because they are concerned about his academic performance. The scheduler noted that they are concerned about his inattentiveness and that he is forgetting to complete his homework. He is a 16-year-old male. Your clinic sent a packet with the Child Behavior Checklist (CBCL), Youth Self Report (YSR), and Teacher Report Form (TRF) to be completed prior to his intake visit. Your clinic has a local prevalence rate (i.e., base rate) of 50%. On the Attention Problems subscale, his mother's rating resulted in a *T*-score of 75, his teacher's in a *T*-score of 68, and John's in a *T*-score of 65. Black ovals indicate positions on the nomogram, and gray ovals indicate prior steps. The combined results move the initial probability of 50% to a posterior probability of 77%. Note that the TRF and YSR scores correspond with only moderate elevations compared to norms and to clinical samples. ADHD = attention-deficit/hyperactivity disorder; DLR = diagnostic likelihood ratios.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

1. Pediatrics AAo. ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Pediatrics.* 2011;128(5):955-965.
2. Pelham WE, Jr., Fabiano GA, Massetti GM. Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *J Clin Child Adolesc Psychol.* 2005;34(3):449-476.
3. Youngstrom EA. Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *J Clin Child Adolesc Psychol.* 2013;42(1):139-159.
4. Achenbach TM, Rescorla LA. *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2001.
5. Biederman J, Monuteaux MC, Kendrick E, Klein KL, Faraone SV. The CBCL as a screen for psychiatric comorbidity in paediatric patients with ADHD. *Arch Dis Child.* 2005;90(10):1010-1015.
6. Chen WJ, Faraone SV, Biederman J, Tsuang MT. Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *J Consult Clin Psychol.* 1994;62(5):1017-1025.
7. Crystal DS, Ostrander R, Chen RS, August GJ. Multimethod assessment of psychopathology among DSM-IV subtypes of children with attention-deficit/hyperactivity disorder: self-, parent, and teacher reports. *J Abnorm Child Psychol.* 2001;29(3):189-205.
8. Derks EM, Hudziak JJ, Dolan CV, Ferdinand RF, Boomsma DI. The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Compr Psychiatry.* 2006;47(2):116-122.
9. Doyle A, Ostrander R, Skare S, Crosby RD, August GJ. Convergent and criterion-related validity of the Behavior Assessment System for Children-Parent Rating Scale. *J Clin Child Psychol.* 1997;26(3):276-284.
10. Hudziak JJ, Copeland W, Stanger C, Wadsworth M. Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: A receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry.* 2004;45(7):1299-1307.
11. Tripp G, Schaughency EA, Clarke B. Parent and teacher rating scales in the evaluation of attention-deficit hyperactivity disorder: contribution to diagnosis and differential diagnosis in clinically referred children. *J Dev Behav Pediatr.* 2006;27(3):209-218.
12. Vaughn AJ, Hoza B. The incremental utility of behavioral rating scales and a structured diagnostic interview in the assessment of attention-deficit/hyperactivity disorder. *Journal of Emotional and Behavioral Disorders.* 2013;21(4):227-239.
13. Aebi M, Winkler Metzke C, Steinhausen HC. Accuracy of the DSM-oriented attention problem scale of the child behavior checklist in diagnosing attention-deficit hyperactivity disorder. *Journal of Attention Disorders.* 2010;13(5):454-463.
14. Gray GE. *Evidence-Based Psychiatry.* 1st ed. Washington, D.C.: American Psychiatric Publishing, Inc.; 2004.
15. Frazier TW, Youngstrom EA. Evidence-based assessment of attention-deficit/hyperactivity disorder: using multiple sources of information. *J Am Acad Child Adolesc Psychiatry.* 2006;45(5):614-620.

ACCEPTED MANUSCRIPT

16. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *J Pediatr Psychol.* 2014;39(2):204-221.
17. 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.
18. American Psychiatric Association. Diagnostic and statistical manual of mental disorders : DSM-5. Arlington, VA: APA; 2013.
19. Cohen P, Cohen J, Kasen S, et al. An epidemiological study of disorders in late childhood and adolescence--I. Age- and gender-specific prevalence. *Child Psychology and Psychiatry & Allied Disciplines.* 1993;34(6):851-867.
20. Cuffe SP, Moore CG, McKeown RE. Prevalence and correlates of ADHD symptoms in the national health interview survey. *Journal of Attention Disorders.* 2005;9(2):392-401.
21. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *The American Journal of Psychiatry.* 2007;164(6):942-948.
22. Brown LK, Hadley W, Stewart A, et al. Psychiatric disorders and sexual risk among adolescents in mental health treatment. *J Consult Clin Psychol.* 2010;78(4):590-597.
23. Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *Int J Methods Psychiatr Res.* 2009;18(3):169-184.
24. Milich R, Licht BG, Murphy DA, Pelham WE. Attention-deficit hyperactivity disordered boys' evaluations of and attributions for task performance on medication versus placebo. *J Abnorm Psychol.* 1989;98(3):280-284.
25. Whalen CK, Henker B, Hinshaw SP, Heller T, Huber-Dressler A. Messages of medication: Effects of actual versus informed medication status on hyperactive boys' expectancies and self-evaluations. *J Consult Clin Psychol.* 1991;59(4):602-606.
26. Carlson CL, Pelham WE, Milich R, Hoza B. ADHD boys' performance and attributions following success and failure: Drug effects and individual differences. *Cognit Ther Res.* 1993;17(3):269-287.
27. Hart EL, Lahey BB, Loeber R, Hanson KS. Criterion validity of informants in the diagnosis of disruptive behavior disorders in children: a preliminary study. *J Consult Clin Psychol.* 1994;62(2):410-414.
28. Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *J Abnorm Child Psychol.* 1991;19(1):75-95.
29. Sibley MH, Pelham WE, Jr., Molina BS, et al. Diagnosing ADHD in adolescence. *J Consult Clin Psychol.* 2012;80(1):139-150.
30. Kaufman J, Birmaher B, Brent D, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry.* 1997;36:980-8.
31. Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? *Compr Psychiatry.* 1983;24(5):399-411.
32. Association AP. *Diagnostic and Statistical Manual of Mental Disorders* Fourth Ed., Text Revision ed. Washington D.C.: American Psychiatric Association; 2000.

ACCEPTED MANUSCRIPT

33. Youngstrom EA, Meyers OI, Demeter C, et al. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders.* 2005;7(6):507-517.
34. van Buuren S, Oudshoorn CGM. *Multivariate Impulation by Chained Equations.* Leiden, The Netherlands: TNO; 2000.
35. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* 1988;240(4857):1285-1293.
36. Venkatraman ES, Begg CB. A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. *Biometrika.* 1996;83(4):835-848.
37. Venkatraman ES. A permutation test to compare receiver operating characteristic curves. *Biometrics.* 2000;56(4):1134-1138.
38. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.
39. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA.* 1994;271(9):703-707.
40. Algorta GP, Dodd AL, Stringaris A, Youngstrom EA. Diagnostic efficiency of the SDQ for parents to identify ADHD in the UK: a ROC analysis. *Eur Child Adolesc Psychiatry.* 2016;25:949-57.
41. Abikoff H, Courtney M, Pelham WE, Jr., Koplewicz HS. Teachers' ratings of disruptive behaviors: the influence of halo effects. *J Abnorm Child Psychol.* 1993;21(5):519-533.
42. De Los Reyes A, Augenstein TM, Wang M, et al. The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychol Bull.* 2015;141:858-900.
43. Bossuyt PM, Reitsma JB. The STARD initiative. *Lancet.* 2003;361(9351):71.
44. Jarrett MA, Van Meter AR, Youngstrom EA, Hilton DC, Ollendick TH. Evidence-based assessment of ADHD in youth using a receiver operating characteristic (ROC) approach. *J Clin Child Adolesc Psychol.* 2016 Oct 24:1-13. [Epub ahead of print]
45. DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. *ADHD Rating Scales-IV: Checklists, Norms and Clinical Interpretation.* New York: Guilford; 1998.
46. Collett BR, Ohan JL, Myers KM. Ten-year review of rating scales. V: scales assessing attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 2003;42(9):1015-1037.
47. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Evidence base of clinical diagnosis: Designing studies to ensure that estimates of test accuracy are transferable. *British Medical Journal.* 2002;324:669-671.

ACCEPTED MANUSCRIPT

Accuracy of Achenbach Scales in the Screening of ADHD in a Community Mental Health Clinic

Joseph S. Raiker, PhD, Andrew J. Freeman, PhD, Guillermo Perez-Algorta, PhD, Thomas W. Frazier, PhD, Robert L. Findling, MD, MBA, Eric A. Youngstrom, PhD

Funding: The work was supported in part by National Institute of Mental Health Grant NIHR01 MH066647 (Principal Investigator, Eric A. Youngstrom).

Acknowledgments: The authors thank the families who participated in this research.

Disclosures:

Dr. Raiker has received research support from the Brain and Behavior Research Foundation, the Children's Trust, NIMH, and NSF.

Dr. Frazier has received research support from, acted as a consultant for, received travel support and/or speaker's honorarium from the Cole Family Research Fund, Simons Foundation, Ingalls Foundation, Forest Laboratories, Ecoeos, IntegraGen, Kugona LLC, Shire Development, Ohio Third Frontier, Bristol-Myers Squibb, NIH, and the Brain and Behavior Research Foundation.

Dr. Findling has received research support from, acted as a consultant for, and/or served on a speaker's bureau for Akili, Alcobra, American Academy of Child and Adolescent Psychiatry, American Psychiatric Press, Bracket, Epharma Solutions, Forest, Genentech, Guilford Press, Ironshore, Johns Hopkins University Press, KemPharm, Lundbeck, Medgenics, Merck, NIH, Neurim, PCORI, Pfizer, Physicians Postgraduate Press, Purdue, Roche, Sage, Shire, Sunovion, Supernus Pharmaceuticals, Syneurx, Takeda, Teva, Tris, Validus, and WebMD.

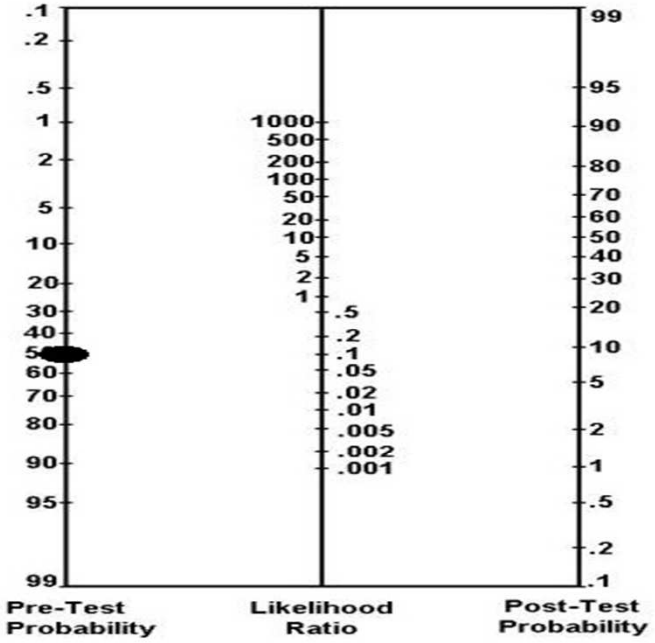
ACCEPTED MANUSCRIPT

Dr. Youngstrom has received grant support from the NIMH, the Society for Clinical Child and Adolescent Psychology, the American Psychological Association, and the Association for Psychological Science. He has served as a consultant to Pearson Publishing, Joe Startup Technologies, Janssen, Lundbeck, and Western Psychological Services about psychological assessment.

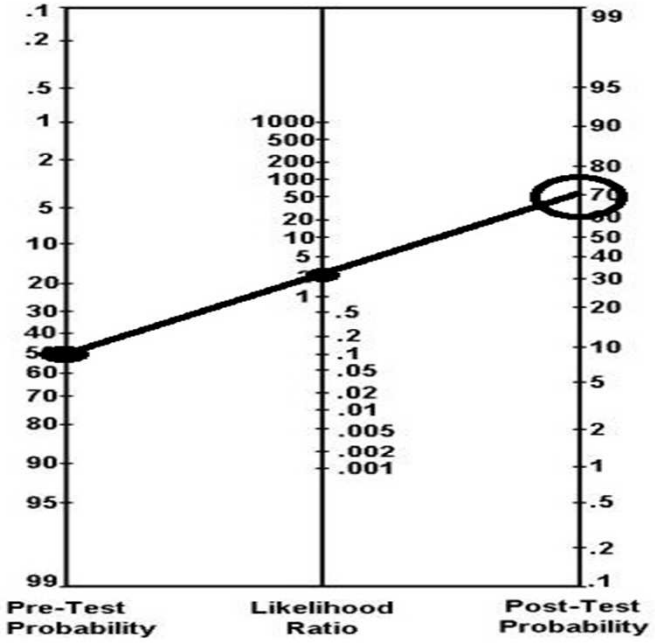
Drs. Freeman and Perez-Algorta report no biomedical financial interests or potential conflicts of interest.

ACCEPTED MANUSCRIPT

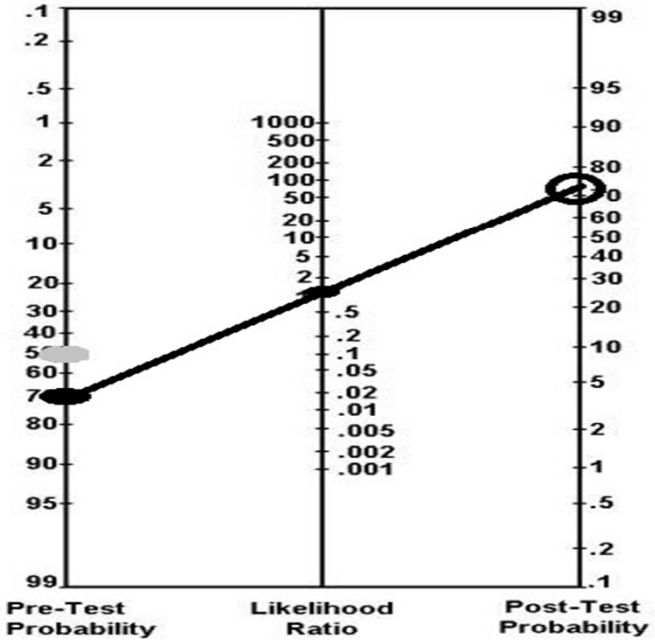
Step 1: Local prevalence of ADHD is 50%



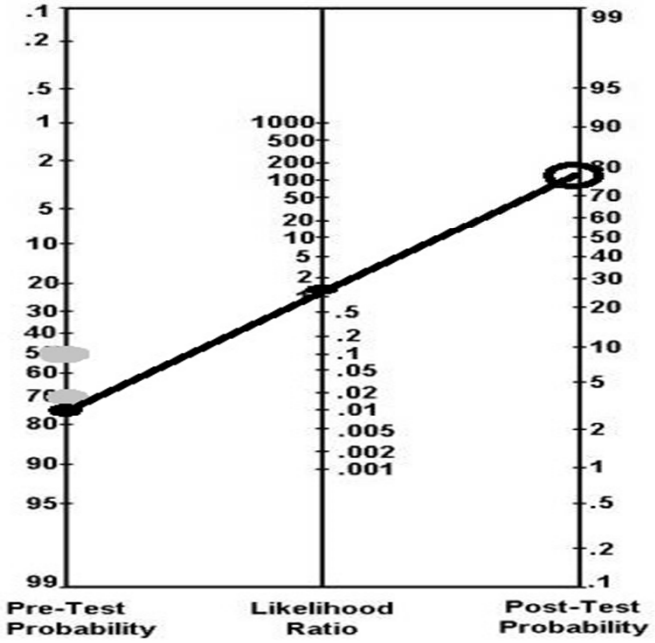
Step 2: CBCL DLR = 2.22



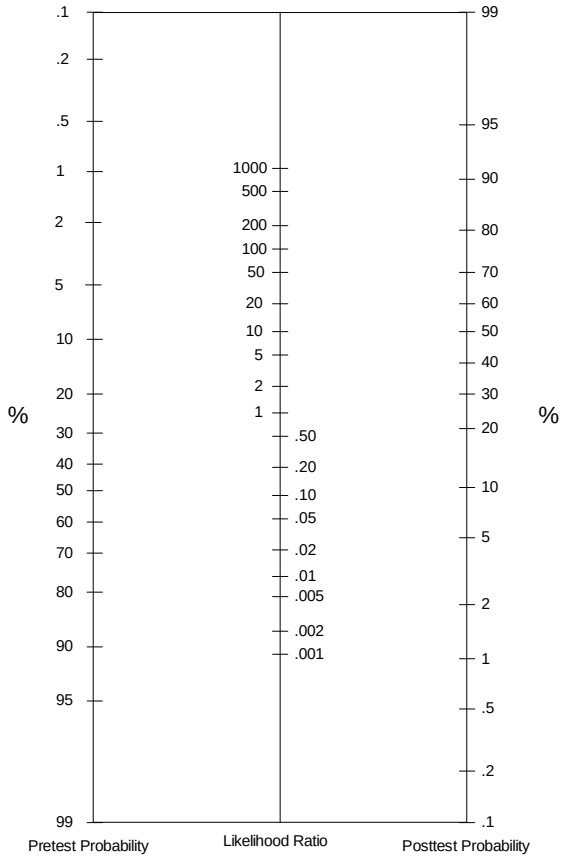
Step 3: TRF DLR = 1.21

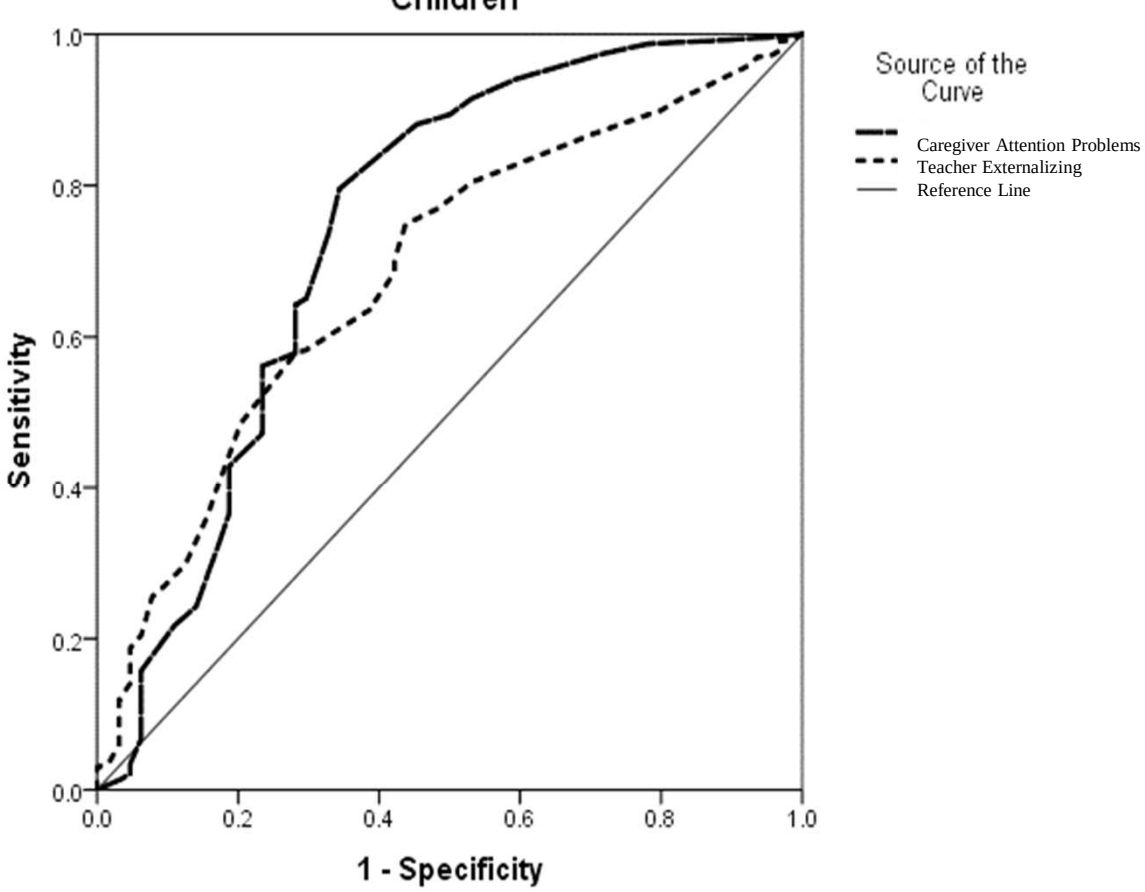


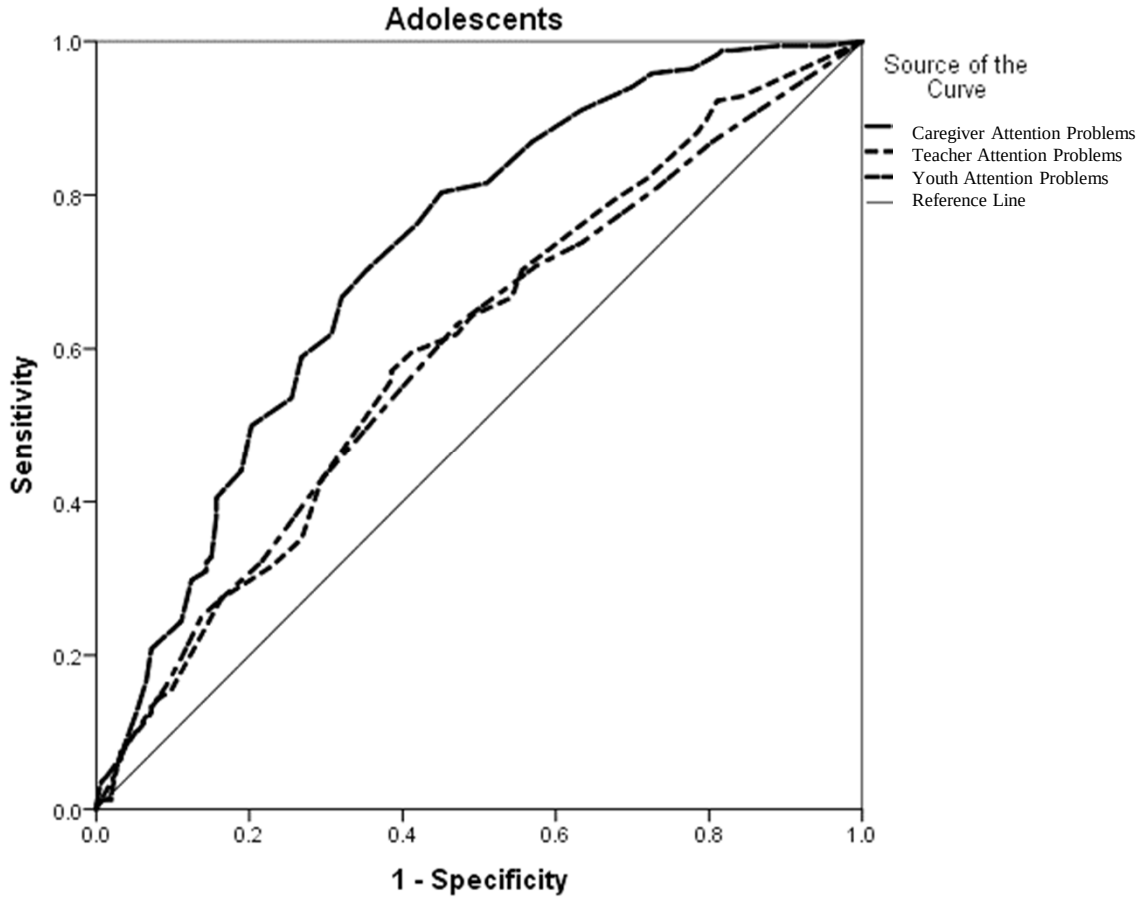
Step 4: YSR DLR = 1.28



ACCEPTED MANUSCRIPT




Children
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Source of the Curve
Caregiver Attention Problems
Teacher Externalizing
Reference Line


Adolescents
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Source of the Curve
Caregiver Attention Problems
Teacher Attention Problems
Youth Attention Problems
Reference Line